mgm . . . for howie, claude and lari
im . . . for wolfgang settekorn
sc . . . for sjb
mjs . . . for christopher mckinstry

Table of contents

Contributors

Leonard Talmy
State University of New York, Buffalo

Raymond Gibbs
University of California, Santa Cruz

Irene Mittelberg
European University Viadrina

Thomas Farmer
Cornell University

Linda R. Waugh
University of Arizona

Mónica González-Márquez
Cornell University

Raymond B. Becker
University of California, Merced

James Cutting
Cornell University

Rafael Núñez
University of California, San Diego

Bonnie Fonseca-Greber
Bowling Green State University

Caroline Vickers
California State University
San Bernardino

Betil Eröz
Middle East Technical University

Stefan Grondelaers
Université Libre de Bruxelles &
Radboud University Nijmeyen

Dirk Geeraerts
University of Leuven

Dirk Speelman
University of Leuven

Sherman Wilcox
University of New Mexico

Jill Morford
University of New Mexico

Eve Sweetser
University of California, Berkeley

Laura Carlson
University of Notre Dame

Patrick L. Hill
University of Notre Dame

Ben Bergen
University of Hawai'i, Manoa

Uri Hasson
The University of Chicago

Rachel Giora
Tel Aviv University

Daniel C. Richardson
University of Reading

Rick Dale
University of Memphis

Michael Spivey
Cornell University &
University of California, Merced

Amanda Brandone
University of Michigan

Roberta Michnick Golinkoff
University of Delaware

Rachel Pulverman
University of Michigan

Mandy J. Maguire
University of Texas at Dallas

Kathy Hirsh-Pasek
Temple University

Shannon M. Pruden
Temple University

Kira Gor
University of Maryland

Seana Coulson
University of California, San Diego

Shimon Edelman
Cornell University

Acknowledgements

Putting together a book like this requires inspiration, patience and lots of help! First of all, we want to thank all of the contributors to this volume. Writing pedagogical material is, without a doubt, one of the most challenging tasks in our profession. How does one explain clearly to a neonate in the field what has become common sense to us as researchers? Not an easy question to answer, yet one to which all of our authors have responded admirably.

The foundation for this volume came from the energy of all of the participants of the first Empirical Methods in Cognitive Linguistics (EMCL) workshop, held at Cornell University, NY, 2003. A warm thank you to you all. The faculty presenters were: Leonard Talmy, Tatiana Chernigovskaya, Seana Coulson, Shimon Edelman, Raymond Gibbs, Kira Gor, Barbara Lewandowska-Tomaszczyk, Scott Liddell, Teenie Matlock, Wolfgang Settekorn, Lera Boroditsky, Chris Sinha, Michael Spivey, and Eve Sweetser. The faculty participants were: Ben Bergen, Claire Cardie, Marianella Casasola, Roberta Golinkoff, Jeff Hancock, Naomi Khoda, Ed Kako, Suzanne Kemmer, Andrej Kibrik, Ken McRae, Rafael Nuñez, Yasuhiro Shirai, Guy Van Orden, Morten Christiansen, Herb Colston, James Cutting, and Joerg Zinken. The student participants were: Carey Benon, Michele Bishop, Sean Flanagan, Uri Hasson, Daniel Richardson, David Havas, Michael Klieman, Chris Koops, Sarah Kriz, Kristin Mulrooney, Andriy M'yachykov, Shweta Narayan, Florencia Reali, Fey Parril, Rachel Pulverman, Sebastian Ross-Hagebaum, Monica Sanaphre, Nathaniel Smith, Bob Williams, Ying Choon Wu, Reka Beneczes, Vito Evola, Maurizio Gagliano, Olga Khomitsevich, Daniel Kislyuk, Sarah Lee, Vera Malioushenkova, Bo Pederson, Sophia Skoufaki, Cristina Soriano, Tatiana Svistounova, Sylvia Tufvesson, Eric Breck and Stanka Fitneva.

For financial support we want to thank the Cognitive Studies Program at Cornell University and the National Science Foundation, in particular Guy Van Orden.

A heart-felt thank you from mgm to her advisors, Michael Spivey, Seana Coulson and James Cutting for allowing her to undertake such an ambitious venture.

A huge thank you to James Cutting, editor of Psychological Science, for allowing mgm to plague him with endless editorial concerns. (Famous last words: this will only take a minute!)

Shimon Edelman's (past chair of the Cognitive Studies Program at Cornell University) adventurous spirit demonstrated by encouraging a second year graduate student in word and deed to begin what have become the international Empirical Methods in Cognitive Linguistics workshops, deserves a generous round of applause. If he hadn't said yes, it is unlikely we would have come this far.

We also want to thank Johannes Wagner and Anders Hougaard for organizing the inspiring Language and Communication workshop, where mgm met her dear friend Uri Hasson, and after much discussion realized that EMCL needed to happen.

The support staff at Cornell was absolutely invaluable in getting so many things done, both with the workshops as with this volume. A tremendous thank you to Linda LeVan, Elizabeth Chandler, Pam Cunningham, Mary Lou DeBoer, Keith Daniels, and Fred Horan.

Thanks are also due to the co-organizers of the two subsequent EMCL workshops. In Portsmouth, England, a thank you to Chris Sinha for inviting the workshop to come to his Language, Culture and Mind Conference (2004), and to Stephanie Pourcel, Stanka Fitneva and Joerg Zinken, co-organizers with mgm, whose dedication led to a workshop where questions key to the further development of the discipline were brought to the fore. Currently, Javier Valenzuela and mgm are co-organizing EMCL 3 in Murcia, Spain. A thank you to Javier and to the Spanish Cognitive Linguistics Association (AELCO) for inviting us to come!

Thank you to Kees Vaes and the editorial team at John Benjamins for their tremendous patience with this volume, and to the head editors of Human Cognitive Processing, the book series this volume will be a part of, Jan Van Nuyts, Ray Gibbs, and Marcelo Dascal.

A special thank you to Raymond Becker for loving, supporting, and generally taking care of mgm.

Finally, many thanks to our two anonymous reviewers at John Benjamin's and to our student/faculty reviewers, Monica Sanaphre, Raymond B. Becker, Javier Valenzuela, Joseph Hilferty, Uri Hasson, Christina Soriano, James Cutting, Rick Dale, Shweta Narayan, Sebastian Ross-Hagemon, Rick Dale, Nathaniel Smith, Bo Pederson, Fey Parrill, Stanka Fitneva, Erin Hannon, Joel Snyder, Florencia Reali, Benjamin Hiles, Daniel Graham, Vito Evola, Reka Benczes, Daniel Kislyuk, Thomas Farmer, Sophia Skoufaki, and Christian Koops.

Original EMCL committee:
Monica Gonzalez-Marquez, Chair
Seana Coulson
Rick Dale
Irene Mittelberg
Michael Spivey

Introduction

The many faces of research in Cognitive Linguistics

The Editors

1. Introduction

Cognitive Linguistics (CL) has been asking theoretical questions for a long time. Most have revolved around understanding how language fits in with everything else about us: the ways we are in our environments, the ways we feel, and the ways we imagine. Although CL is fundamentally committed to the psychological reality of its theoretical constructs, much of this work is based purely on the linguistic intuitions of the theoreticians. However, there is a growing awareness that linguistic theory should be grounded in the observation of language usage, in experimental tests of its validity, and in general knowledge of cognitive function. Work in cognitive metaphor, cognitive grammar, psycholinguistics, discourse management, conceptual integration, and spatial cognition has produced tantalizing proposals about the conceptual underpinnings of human languages. A major movement in cognitive linguistics has thus developed around the commitment to pursue empirical studies that might help substantiate its claims, and to develop a coherent account of the connection between language and cognition.

The Empirical Methods in Cognitive Linguistics (EMCL) workshops were conceptualized as a practical bridge between theoretical and empirical work. Because Cognitive Linguistics does not assume isolation for linguistic processes, most of the methods used to study cognition can potentially be adapted to investigating language. At the same time, language usage data, including co-speech gesture and signed languages, are sources of overt manifestations of cognitive processing, situated cognition, and social behavior. The potential for methodological cross-fertilization is already obvious in the broad range of methods in use. The latest International Cognitive Linguistics Conferences have yielded provocative offerings on everything from eye-tracking studies, to corpus analyses, to non-linguistic behavioral studies, to computational modeling.

Though in principle the methodological variety is a fortuitous event, in practice, the prospect of an empirical investigation can be overwhelming, especially to a novice researcher trained in a more rationalist mode of inquiry. Therefore, in lieu of publishing a sort of proceedings for EMCL, this volume is intended as a handbook to exploring the empirical dimension of the theoretical questions raised by CL. It presents the reader with guidelines for employing methods from a variety of intersecting disciplines, laying out different ways of gathering empirical evidence for claims originally motivated by theoretical considerations or by a particular set of data. Essentially, the authors not only introduce methods that are well-established in their respective fields and provide examples of their application, they also share their experience in working with these methods and alert the reader to both their promise and shortcomings. Our hope is that these chapters will be a resource of basic facts a linguist needs to know before tackling a new methodology and will be a useful reference later on.

2. Volume organization

The volume is divided into 5 sections. They are *Methods and Motivations, Corpus and Discourse Analysis, Sign Language and Gesture, Behavioral Research* and *Neural Approaches.*

2.1 Methods and motivations

2.1.1 *Why cognitive linguists should care more about empirical methods by Raymond Gibbs*

Gibbs describes several ways that cognitive linguists and cognitive psychologists can more fruitfully explore their respective claims about the interaction of thought, language, and the body in their empirical work. He outlines exactly what is empirical about cognitive linguistics, how this work may be relevant to psychological theories of mind and language, and also how cognitive linguists need to alter their methods and their theories to better meet the demands of a scientific discipline. A special focus is placed on the need for greater reliability in cognitive linguistic methods of linguistic analyses, and on the importance of falsification in cognitive linguistic theory construction.

2.1.2 *They actually said that? An introduction to working with usage data through discourse and corpus analysis by Irene Mittelberg, Thomas Farmer & Linda R. Waugh*

The chapter by Mittelberg, Farmer, and Waugh serves as an introduction to the more specific chapters on discourse and corpus analysis. It provides the reader with some selected strategies to investigate cognitive and linguistic structure grounded in situated language use, including talk-in-interaction, newspaper discourse, co-speech gesture, and signed languages. Furthermore, corpus-based research is discussed focusing on methods for extracting and examining naturally-occurring language patterns. The chapter also contains a summary of commonly used terminology as well as an annotated list of commonly used corpora. The authors hope to inspire the reader to consider discourse and corpus analysis as one of the keys to a fuller understanding of human cognition and social behavior.

2.1.3 *An introduction to experimental methods for language researchers by Mónica González-Márquez, Raymond B. Becker & James Cutting*

The chapter by Gonzalez-Marquez, Becker and Cutting aims to introduce both established researchers and graduate students without experimental research experience to the general principles of laboratory methodology employed to investigate cognition. The goals of the chapter are to demystify the research process, as well as to help the novice researcher understand how to use published research as a guide for their own work. The chapter describes basic experimental design, explanations of the null and experimental hypotheses, and the use of control studies, amongst other key topics. The chapter ends by discussing the growing importance of cultural and linguistic differences in cognitive psychology.

2.1.4 *Inferential statistics in the context of empirical cognitive linguistics by Rafael Núñez*

In the last chapter to this section, Núñez completes the introductory survey by analyzing some of the methodological difficulties involved in combining evidence from psychology and linguistics. Inferential statistics is described as a type of scientific tool for treating empirical data. The reader is then walked through the most common statistical methods (parametric and nonparametric) used to analyze cognitive data.

2.2 Corpus and discourse analysis

2.2.1 *Multiple empirical paths to a complex analysis of discourse by Linda R. Waugh, Bonnie Fonseca-Greber, Caroline Vickers & Betil Eröz*

The chapter by Waugh, Vickers, Fonseca-Greber & Eroz exemplifies the benefits of an integrated approach to the study of discourse that examines culturally contextualized small corpora of authentic language use with a rich, fine grained analysis derived from various empirical approaches. The topics addressed include changes in the forms, meanings and use of French pronouns and how an analysis of the use of indefinite pronouns (like 'one') shows how participants index their linguistic and cultural identity; how native and nonnative speakers of English in a course together accommodate to each other as they try to perform a task and instances in which they successfully come to work within a shared interpretive frame, i.e., ways of knowing and of experiencing the world; and how the classroom behavior of international students is different from Americans' and how their ways of interacting are related to classroom and cultural patterns in their home country.

2.2.2 *Case for a cognitive corpus linguistics by Stefan Grondelaers, Dirk Geeraerts & Dirk Speelman*

Grondelaers, Geeraerts, & Speelman introduce the field of corpus research by briefly presenting its three main areas: corpus design, information retrieval, and data analysis. All three areas are dealt with by example to increase accessibility, and include pointers to important resources (like corpus tools). In addition, corpus-based linguistic research is discussed from a slightly broader methodological perspective, viz. the methodological relationship between corpus analysis and other empirical methods (specifically, experimental research), i.e. what type of linguistic problems can be dealt with using corpora. The natural complementarity between corpus research and experimentation is brought to the fore.

2.3 Sign language and gesture

2.3.1 *Empirical methods in signed language research by Sherman Wilcox & Jill Morford*

Wilcox & Morford explore various empirical approaches to the study of signed languages, describing the benefits and pitfalls of each. They discuss the use of historical data from several sources including documents and films, methods for eliciting data, and conversational data collected from various sources. Particular attention is given to issues involved with collecting video data, including logistical questions such as setting, data collection, and confidentiality. In addition, they address issues related to variability within the signed language community and modality. Their chapter ends with a discussion of alternative methodologies for conducting empirical studies.

2.3.2 *Looking at space to study mental spaces: Co-speech gesture as a crucial data source in cognitive linguistics by Eve Sweetser*

Sweetser begins by laying out some of the well-established findings underlying modern gesture research, for example, the evidence that gesture and speech are co-produced as a single neural package, and that a specific pattern of neural packaging is shared by a given language community. She goes on to introduce some of the basic frameworks of gesture analysis and evaluates the ways in which they are useful to cognitive linguists. She examines how these tools can help us to further understand ongoing language and cognition while observed gestures are performed – and vice versa. Sweetser argues in particular that the tools of Mental Spaces Theory are extremely productive in this endeavor. Her intent is to help readers with differing research goals learn to make use of gestural data – and to give them realistic expectations as to the kinds of work involved and the kinds of results that may emerge.

2.3.3 *Methodology for multimodality: One way of working with speech and gesture data by Irene Mittelberg*

The chapter by Mittelberg provides an example of the kinds of decisions and methodological steps that shape empirical work with multimodal usage data consisting of spoken discourse and its accompanying gestures. Pointing to ways in which co-speech gesture not only engages the body of the speaker, but also the eye of the linguist, the author walks the reader through the process of recording, transcribing, coding, and annotating speech and gesture data. Additional approaches of particular interest within cognitive-functionalist linguistics are also discussed.

2.4 Behavioral experiments

2.4.1 *Experimental methods for studying language and space by Laura Carlson & Patrick L. Hill*

Carlson & Hill cover a diverse set of empirical methods used to investigate spatial language. These include placement tasks, acceptability judgments, speeded verification, and production. For each method, they begin by offering a description of the method, and present examples of research that have used it. They discuss the strengths of each method, focusing on the kinds of conclusions that it licenses as well as its respective weaknesses, focusing on the potential problems that exist, and addressing how a researcher can confront these limitations. Examples from the literature are used to illustrate. The last section of the chapter argues that a combination of these methods is important in offering converging evidence.

2.4.2 *Experimental methods for simulation semantics by Ben Bergen*

Bergen's chapter discusses the leading methods for investigating perceptual and motor imagery during language understanding. The chapter is divided into two main methodological sections, motor imagery and perceptual imagery. The methods covered include interference effects, such as the Action-sentence Compatibility Effect or ACE, the Perky effect, cross-modal interference, and length-of-simulation measures. Together, this collection of methods constitutes a sound basis for the behavioral study of mental simulation as it participates in natural language understanding.

2.4.3 *Experimental methods for studying the mental representation of language by Uri Hasson & Rachel Giora*

Hasson and Giora present research methods that have been employed to study the interpretation of linguistic structures, i.e., the mental representation which is the product of the comprehension process. The chapter's two goals are to present these research methods and to expose the reader to relevant findings disclosed by the use of such methods. The chapter discusses methods such as online reading-time measures. priming methodologies, memory measures, and self-report measures. The technical aspects of each procedure and its theoretical strengths and weaknesses are also covered. Finally, the authors review the theoretical constructs and assumptions underlying each of the procedures, as these ultimately determine the interpretation of the data.

2.4.4 *Eye movements in language and cognition: A brief introduction by Daniel C. Richardson, Rick Dale & Michael Spivey*

A number of different methods for eye-tracking, and experimental applications of eye-tracking, have surfaced throughout the last few decades to complement more conventional reaction-time and accuracy measures of processing. This chapter provides a methodological tutorial limited to eye-tracking methods in language research. It describes how patterns and timing of saccades and fixations have been used to test theories of written word recognition, sentence processing and, given a highly relevant visual context, even spoken language comprehension. Later, it goes on to discuss research where eye movements have proven informative for spoken language comprehension, mental models, and semantic memory, in situations where the visual display is completely blank. This latter collection of findings points to the spatial structure underlying language, possibly especially accessible through the use of eye-tracking.

2.4.5 *Speaking for the wordless: Methods for studying the foundations of* Cognitive linguistics in infants
by Amanda Brandone, Roberta Michnick Golinkoff, Rachel Pulverman, Mandy J. Maguire, Kathy Hirsh-Pasek & Shannon M. Pruden

The authors describe the most useful methods to studying the incipient knowledge non-verbal infants have of the events and actions they observe. Habituation and intermodal preferential looking paradigms are discussed in detail. In the habituation paradigm, researchers utilize infants' preference for novelty. Infants are repeatedly shown a stimulus until their visual fixation to the display has decreased by a predetermined criterion. Novel test stimuli are then introduced with the expectation that if infants notice and find the changes to be interesting, they will watch the novel displays longer than the familiar ones. In the intermodal preferential looking paradigm (IPLP), infants are presented with a split-screen video presentation and acentral audio. The premise is that if children understand what they are hearing, they should spend more time looking to the display that matches the audio. After illustrating these methods, the authors describe research findings that inform the cognitive linguistics program.

2.4.6 *Experimental* Study of first and second language morphological processing
by Kira Gor

This chapter discusses the experimental paradigm used in research on the processing of inflectional morphology by first and second language speakers. It mainly focuses on the available data on regular and irregular verb processing in English and several languages with complex inflectional morphology, such as Russian, Italian, German, Norwegian, and Icelandic. The continuing debates between the proponents of the dual- and single-system approaches and some "hybrid" theories center around the claim that symbolic rule processing cannot be influenced by input-based linguistic probabilities or phonological similarity to the existing words belonging to the "neighbors"' and "enemies"' clusters. Accordingly, the paper addresses the ways in which the role of linguistic probabilities can be studied experimentally. It reviews the elicitation techniques aimed at triggering real and nonce verb generation, and how by controlling the input to second language learners, experimenters are able to compare the native and non-native mechanisms of morphological processing.

2.5 Neural approaches

2.5.1 *Electrifying results: ERP data and cognitive linguistics by Seana Coulson*

Event-related brain potentials (ERPs) represent electrical activity in the brain that is time-locked to the onset of a cognitive or motor event. A direct index of brain activity with high temporal resolution, ERPs are known to be sensitive to many of the processing operations involved in the production and comprehension of language. But what are ERPs, really? And, what are they good for? This chapter reviews methods and data in the domain of electrophysiology of language comprehension. It begins with a general description of the electroencephalogram and event-related brain potentials (ERPs) and give an overview of language-sensitive ERP components such as the N400. Discussion of various ERP findings in psycholinguistics is intended to highlight how ERPs can be used to address questions about the representation and timing of cognitive processes, and how electrophysiological data can be used to complement experimental findings using behavioral paradigms. Finally, the author suggests how ERPs might be used to experimentally address issues in cognitive linguistics. It then describes the constraints one needs to consider in designing an ERP experiment, and gives examples of research which does and does not employ the strengths of the technique.

2.5.2 *Bridging language with the rest of cognition: Computational, algorithmic and neurobiological issues and methods by Shimon Edelman*

The computational program for theoretical neuroscience proposed by Marr and Poggio calls for a study of biological information processing on several distinct levels of abstraction. At each of these levels – computational (defining the problems and considering possible solutions), algorithmic (specifying the sequence of operations leading to a solution) and implementational – significant progress has been made in the understanding of cognition. In the past three decades, computational principles have been discovered that are common to a wide range of functions in perception (vision, hearing, olfaction) and action (motor control). More recently, these principles have been applied to the analysis of cognitive tasks that require dealing with structured information, such as visual scene understanding and analogical reasoning. Insofar as language relies on cognition-general principles and mechanisms, it should be possible to capitalize on the recent advances in the computational study of cognition by extending its methods to linguistics.

Foreword

Leonard Talmy

1. Introduction

The new insights into the system of conceptual structuring in language that have been coming from the relatively recent tradition of cognitive linguistics have rested mainly on the methodologies already standard in the field of linguistics overall: introspection in conjunction with theoretical analysis. The aim of the workshop that the present volume arises from was to help foster the application of additional methodologies to this emerging body of understanding. The spirit of the workshop and the papers here has been to value all of the applicable methodologies for their distinctive contribution to the total picture. Each methodology can be seen as having certain capacities and limitations that accord it a particular perspective on the nature of conceptual organization in language. In this respect, no single methodology is privileged over others or considered the gold standard of investigation.

Though not all of them were represented at the workshop or are in this volume, the range of methodologies that apply to conceptual structure in language includes the following: introspection into the meanings and structures of linguistic forms and expressions, whether in isolation or in context, as well as the comparison of one's own introspections with those reported by others (the more recent notion of "meta cognition" largely overlaps with that of introspection); the comparison of linguistic characteristics across typologically distinct languages and modalities (e.g., spoken and signed language); the examination of how speech events interact with context, such as with the physical surroundings, the participants' background knowledge, or the cultural pattern; the analysis of audiovisual recordings of naturally occurring communication events, including their text, vocal dynamics, gesture, and body language; the (computer-aided) examination of collated corpora, often annotated; the examination of cumulatively recorded observations of linguistic behavior, as by children acquiring language; the experimental techniques of psycholinguistics; the instrumental probes of the brain's linguistic functioning in neuroscience; and the simulations of human linguistic behavior in artificial intelligence. Used in conjunction with all of these is the methodology of analytic thought, which includes the systematic manipulation of ideas, abstraction, comparison, and reasoning, and which is itself introspective in character, though with its object of attention not limited to language, as in the case of the linguistic introspection otherwise treated here. A selection of these methodologies is considered next for their respective capacities and limitations, so as to demonstrate their complementary character.

2. Introspection

In addition, introspection exhibits a particular profile of both capacities and limitations in accessing different aspects of language, as described below, and it is its pattern of limitations that has in part spurred the use of certain other methodologies to fill in for them. The methodology of introspection begins this account and occupies some space because it has been central in the development of cognitive linguistics and continues as its main methodology, and because its particular profile of limitations has in part led to the pattern in the use of other methodologies. Linguistic introspection is conscious attention directed by a language user to particular aspects of language as manifest in her own cognition. More specifically, certain aspects of language spontaneously or through evocation can appear in a language user's consciousness – what can here be termed "first-level consciousness". In the same language user, a second level of consciousness can also occur that has as its object the contents of the first level of consciousness. This second-level consciousness – or attention – can be volitionally evoked and directed at a chosen linguistic target. Aspects of language differ in their readiness to appear in first-level consciousness. And, if present there, they differ in their amenability to second-level attention. An aspect of language is more amenable if it has greater strength and clarity and can remain more stably present in first-level consciousness while attention is directed at it, whereas it is less amenable if it is fainter, vaguer, or more elusive under such attempted scrutiny. As a cover term spanning such first-level readiness and second-level amenability, aspects of language will here be said to differ in their "accessibility" to consciousness, attention, or introspection (see Talmy forthcoming).

The accessibility of an aspect of language to directed conscious attention depends at least on the following five factors: cognitive organization in general, particulars of an individual's cognition, the current situation, conditions of attending, and the categorial object of attention. For the first factor, cognition across individuals appears to be structured in such a way – whether innately or from common developmental conditions – as to privilege certain aspects of language over others along the accessibility gradient. As a second factor, though, due to individual differences in cognition – whether these result from innate differences or from training or practice – particular aspects of language can be above or below average in their accessibility to consciousness in a language user, thus able to diverge within limits from their usual ranking. As the third factor, the accessibility that various aspects of language afford to consciousness can vary over time within a single individual in accord with changes in the situation or his concerns. An individual might, for example, attend more strongly to the exact wording of a lawyer and more to the tonality of an intimate. Fourth, the accessibility of various aspects of language to consciousness differs in accordance with what can be called the "conditions of attending", three of which can be suggested.

In the first condition, an individual uses second-level consciousness to attend in isolation to some aspect of language within some small excerpted portion of discourse such as a word or sentence. In the second condition, an individual endeavors to attend with second-level consciousness to aspects of language that appear in first-level consciousness in the course of an ongoing communication event that she participates in as speaker or listener. The third condition rests on the assumption that some trace of first-level consciousness of some aspect of language can continue to be present for a short time, gradually fading away, directly following its evocation in some communication event. In some cases and to some degree, an individual can use second-level consciousness to attend to this brief perseveration of first-level consciousness to examine what its contents had been during their immediately preceding activation. The fifth factor, the category of the object of attention, can range from the meaning of a word, through the grammaticality of a sentence, to the intonation of one's speech. This factor is best considered in conjunction with the properties of the prior factors, as in what now follows.

2.1 First condition of attending

To begin with the first condition of attending, human cognition seems organized in such a way that categories of language aspects differ in their accessibility to introspection in isolation. Of the most accessible categories – ones that are strong, clear, and stable as objects of attention in isolation – the foremost is meaning: the conceptual content associated with linguistic representations. Not only is meaning the aspect of language that linguistic introspection is best at, but, in addition, introspection has the advantage over other methodologies in seemingly being the only one able to access it directly. Meaning is a consciousness phenomenon and, if it is to be taken on as a target of research, introspection – itself a process occurring in consciousness – is the relevant instrumentality able to reach its venue. Introspection accesses meaning of several types. One type is the meaning of an individual word. Access tends to be greater for an open-class word than for a closed-class word, and greater for a concrete meaning than for an abstract one. Thus, one can readily attend to the meaning of the open-class concrete word *bucket*, less so to the meaning of the open-class abstract word *relation*, and perhaps still less so to the meaning of the closed-class and abstract words *not* and *with*. Also stable under scrutiny in isolation is the overall meaning of a phrase or sentence. Likewise accessible is the derived meaning of an idiom or figure of speech, such as a metaphor, as distinguished from any literal reading such a form might represent. Comparably, as is often apparent to someone pausing to choose his phrasing carefully in writing, one can attend directly to the appropriateness or good fit of a word with respect to its meaning in a given context (or to its register in a given context as mentioned next).

Somewhat less directly involved with meaning but still highly accessible to attention in isolation is the register of a linguistic form – that is, apart from its actual referent, its character as relatively more colloquial, learned, child-oriented, etc. and, hence, its appropriateness for use in an informal, formal, or child-oriented situation, etc. Finally, now apart from meaning or its context, linguistic cognition seems structured in a way to permit relatively easy conscious access to the grammaticality of a phrase or sentence – that is, its degree of conformity to the standard patterns of syntax and morphology in one's language – though this can vary substantially across individuals, especially as a consequence of training. Most of generative linguistics rests on an assumption of the reliability of such grammaticality judgments.

Still under the condition of attending to them in isolation, some aspects of language offer only moderate rather than strong accessibility to introspection. Returning to meaning as the object of attention, the strong access to the meaning of a word, noted earlier, holds whether the word is composed of one or more morphemes. But in the latter case, the meaning of a bound closed-class morpheme – that is, of an affix or clitic – seems only moderately accessible. Thus, while the meaning of the whole word *unresettable*, as in *This kind of trap is unresettable*, is quite available to direct examination, the meanings of the affixes *un-*, *re-*, and *-able* seem less fully and immediately available without the aid of analytic manipulations. However, this aspect of language may be quite sensitive to differences across individual speakers. Such a difference may have existed between my two main consultants for Atsugewi – a polysynthetic Hokan language of northern California – both with little Western schooling and with their native language unwritten. One consultant could provide the single multi-affixal verb best expressing a situation I had depicted, but could not identify any of the affixes within the verb, such as the particular "Cause prefix" representing the specific cause of the main event. The other consultant could not only provide the best verb for the depicted situation, but could proceed to volunteer systematic variations of cause within the situation and provide the corresponding series of verbs, each one with a different Cause prefix. The second consultant would seem to have had greater conscious access to the morphological structure of words in her language.

A further moderate case is that of access to the different meanings associated with a particular morphemic shape. For example, a person asked to think of the various senses of the noun *stock* might come up with several, but scarcely all, of the following: 'soup base', 'stored supply', 'line of descendants', 'farm animals', 'financial instrument', 'rifle part', 'fragrant flowered plant species'. One could then ask, though, what happened to the remaining senses in the respondent's cognition. On hearing the word *stock* in contexts that evoked them, the respondent would no doubt have come up with those very senses. Apparently, our cognition is organized in a way that allows particular senses of a word to come into consciousness in the relevant contexts, but not as a full connected set under introspection – though introspection does yield a few. The full set can again typically be achieved only with the aid of analytic procedures, like those giving rise to dictionaries through a kind of corpus research. A similar accessibility pattern seems to hold for a comparable aspect of language: approximate synonyms in a lexicon. If asked to think of other words with roughly the same meaning as, say, *tendency*, a respondent might come up with a couple, but probably not all, of the following: *inclination, leaning, disposition, proneness, propensity, proclivity*. If we think of a certain concept, our cognition is apparently so organized as to present in consciousness one or two of the lexical forms that express it and, with introspection, to present several more such forms, but not to present the extended set. As before, the extended set can be achieved only with the aid of analytic procedures through another kind of corpus research, the kind that leads to a thesaurus.

Finally, now somewhat less directly involved with meaning, the lexical category of a word – noun, verb, adjective, etc. – also seems moderately accessible to introspection. Our cognition seems structured to allow us some sense of lexical category, although individuals surely vary on this and, even the most natively gifted individual would need training or practice to be able to articulate this sense.

As the last circumstance under the condition of attending to a linguistic entity in isolation, some aspects of language appear to remain poorly or not at all accessible to introspection, that is, are elusive or absent under direct attention. Often, the means for getting at such aspects is to combine introspection – employing it where its capacities work well – with analytic manipulation. In the area of linguistic meaning, while the overall meaning of a word is one of the aspects of language most accessible to direct attention, the specific semantic components that structurally comprise that meaning largely are poorly accessible. For example, one can reasonably well attend in isolation to the overall meaning of the preposition *across* in its locative sense, as in *The board lay across the road*. And, if asked to, one could probably come up with a vague sense of what seemed like one of its components of meaning, say, 'perpendicularity of two axes'. But the combination of semantic introspection with analytic manipulation can tease out no fewer than nine criterial components of meaning (see Talmy 2003, 2006). The method here is to systematically alter individual elements (especially spatial relations) within the referent situation and use one's semantic introspection to see if the original sentence with *across* still applies to it, or if now a sentence with a different preposition would apply instead. The situational elements needed for *across* to apply must then correspond to components of its meaning. Thus, one can little directly discern the semantic components of a word's meaning through introspection, but one *can* use introspection – in fact, perhaps one must use it – where it functions appropriately, in the experiencing of the overall meaning of a word, as part of a procedure for ferreting out its semantic components.

Another aspect of language that does not simply pop into awareness on introspection is that of syntactic principles and patterns. For example, if asked to consider the two sentences *Whose dog did our cat bite?* and *Whose dog bit our cat?*, an average English speaker would have little direct sense for what it is about the first sentence that (among other syntactic characteristics) requires the inclusion of the word *did*, the basic form of the verb *bite*, and the positioning of this verb at the sentence's end, while the second sentence requires an absence of *did*, the past-tense form of the verb, and the positioning of the verb within the sentence. Again, though, linguists have been able to tease out syntactic regularities governing facts like these by combining their introspective sense for the grammaticality of sentences with the analytic procedure of systematically altering the elements and arrangements within such sentences to uncover patterns.

Finally, there is little if any direct conscious access to the cognitive processes that underlie any of the strongly viable operations of introspection. For example, although we can readily attend to the fact that we get an immediate sense of the meaning of the word *bucket* when we hear it, we cannot attend to the mental processing that led to that meaning coming to mind.

2.2 Second condition of attending

We turn now to the second condition of attending, that is, directing one's attention to various aspects of language occurring while one is engaged as a speaker or listener in ongoing discourse (which could be extended to fluent writing or reading). The various aspects of language present in discourse again seem to range widely in their accessibility to such meta-attention. Thus, strongly accessible are the overall topic or subject matter of some portion of discourse and – though this may well vary across individuals – a sense for the degree of thematic and logical coherence present in the discourse. Also highly accessible is the specific conceptual content expressed by a speaker one is listening to, even down to a relatively fine-grained level if one listens attentively. Perhaps a bit lower in accessibility to meta-attention, though still high, is the conceptual content that one has in mind to express as one speaks, as well as the conceptual content one cues up for one's next turn at speaking while still listening to one's interlocutor.

But certain other aspects of language seem to afford only moderate, little, or no accessibility to direct attention during communication. Thus, during speech, the recombinant core system of language (see Talmy 2004) is accompanied by expressive subsystems that the speaker generally seems able to attend to only sporadically and moderately. Perhaps in order of decreasing accessibility, these include "vocal dynamics", my term for the gradient subsystem in language that includes pitch, loudness, rate, precision, etc.; the speaker's gestures, including not only their unfolding forms but also their timing relative to the ongoing spoken track; and the speaker's facial expressions and body language.

Deviations from well-formedness in speech can attract speaker or hearer attention and can even appear in attention with full salience if they exceed a certain "grace allowance", but those occurring within the grace allowance scarcely attract attention and are only moderately accessible to any attention directed at them. Such deviations include contextually inappropriate forms, misplaced forms, non-optimal choice of words or constructions, grammatical or referential conflict across different sentence portions, self-corrections, incomplete constructions, pauses, inclusion of "uh" and "oh", restarts, interruptions by other speakers, and overlaps with other speakers. Also attracting little attention and perhaps highly affected by any attention directed at it is what a speaker's or hearer's gaze successively lights on, in the former case perhaps influencing what she next says and in the latter case often influenced by what the speaker has just said.

Finally, seemingly inaccessible to conscious attention are the cognitive operations and processes going on that yield the production or the comprehension of speech.

2.3 Third condition of attending

Without examining it in detail, the third condition of attending – directing attention to the memory trace of various aspects of language that have just been manifested – also exhibits different degrees of accessibility. Thus, at the relatively high end of accessibility might be the ability to attend to the thematic topic and conceptual content, still in memory, of some immediately prior discourse. By contrast, the memory of, and hence one's ability to attend to, the exact wording and phrasing that had just been used to represent such conceptual content might be elusive.

2.4 Profile of introspection

A profile has just been sketched of the capacities and limitations that introspection has in accessing aspects of language. Where this profile has peaks and elevations, introspection may be the methodology of necessity or choice. But as its profile dips, other methodologies become increasingly necessary to corroborate or fill in for the introspective findings. And where introspection lacks access altogether, any information about those aspects of language must come from other methodologies.

In addition, researchers can feel a call to other methodologies to the extent that they have either or both of the following two concerns. The first is the concern that observation can affect what is observed. For the present intra-cognitive issue, this is the concern that directing second-level attention at them might disturb the contents of first-level consciousness. The claimed gradient by which first-level contents range from being more stable to being more elusive either can be accepted and cause increasing concern toward the elusive end, or can be questioned even at the proposed stable end. Further, concern is likely to be greater for the second condition of attending than for the first, in part because some of the attention otherwise needed just to maintain adequacy of speaking or listening in a conversation might be diverted to the purpose of self-observation.

The second concern is over the reliability of and efficacy of introspection in the first place. The reservation can be over the existence of consciousness itself or, if that is granted, over whether the contents of consciousness can accurately reflect the unconscious mental processes felt to constitute the bulk of cognitive functioning.

3. Audio- and videographic analysis

Our look at several other methodologies for their respective capacities and limitations can start with that of examining audiovisual recordings of speech events – that is, audio- and videographic analysis. This methodology has certain advantages in accessing aspects of language where introspection is limited. One such advantage is the ability to examine at leisure the online expressive accompaniments of the core speech track that were earlier cited as only of moderate to low accessibility to second-condition introspection. These include a speaker's vocal dynamics, gestures, facial expressions, body language, and gaze direction. Included as well is the exact timing of all these accompaniments, both alone and in relation to each other. Live recording can capture for study many aspects of the core speech track as well – also not readily attended to during the event – such as its deviations from well-formedness. In addition, what a speaker says at any given moment can be regarded as the result of numerous interacting factors, exerting their various calls on the speaker (see Talmy 2000b: Ch. 6). These factors are sensitive to the immediate idea as well as the larger ideational complex needing expression, to assumptions about what the addressee already knows or should be informed of, to what needs emphasis or de-emphasis, etc.; and yields selections as to morphemes and constructions. Introspection cannot readily access all these factors online because there are too many of them, because many are elusive, and because some of them may be out of awareness.

Further, these factors appear to be triggered in their full complement in the cognition of the speaker only when he is engaged in what he understands to be an actual interlocutive event of speaking. Hence, audiovisual recordings register what people actually say during naturalistic speech, which cannot be replicated by the kinds of sentences often constructed out of context for first-condition introspection in isolation.

Finally, longer stretches of live recording permit analysis of the kind of discourse functions central to functional linguistics, which, due to their length, are minimally accessible to the isolative introspection that excels at short linguistic forms.

One main limitation to the methodology of working with spontaneous communication – given that such communication is the result of many interacting cognitive factors – is that it does not lead directly to determining the lawful properties of any single factor by itself. For this, the more isolative, probing methods are generally needed. Perhaps a physical analogy might be that, while one kind of study can characterize the complex pattern in which, say, a feather falls, another kind of study is needed to separate out and individually characterize the contributing components, such as gravity, turbulence, friction, and buoyancy.

4. Corpus analysis

Corpora, whose use constitutes another methodology, share with audiovisual recording a focus on naturalistically produced speech. But they typically capture only its segmental text within their usually written format. Live recording thus has the advantage over corpora in capturing the expressive accompaniments of text and the timing of all these components. The consequent advantage of corpora, though, is to make a large quantity of textual discourse available to searches for particular phenomena of interest.

This advantage is especially great for linguistic phenomena whose frequency of occurrence or range of instantiations is the issue. Thus, corpus research can determine the frequency with which a particular expression is used in various contexts. This can be interpreted as a reflection of its unconsciously registered values along various cognitive parameters. These unconscious values are otherwise reflected in consciousness and available to introspection only approximately – as a phenomenological sense of the expression's colloquiality and register. Searches through chronologically sequenced texts, in addition, can uncover gradual changes in the frequency – and therefore changes in the unconscious parametric values – of an expression, whereas such change is largely too slow for any direct introspection. Frequency assessments can also be made of a particular morpheme's co-occurrence with other morphemes. Once again, introspection could yield an approximate sense for the naturalness of such collocations, but not of their specific patterns of occurrance.

Another special asset of corpus work is that it helps one map out the range of alternative realizations of a linguistic entity. Examples of such ranges are the set of constructions that a particular morpheme can participate in, and the set of polysemous or homonymous senses that a particular morphemic shape can represent. As noted, a dictionary can be regarded as a form of corpus work that addresses this latter example. Direct introspection into such sets of alternatives typically does not come up with the entire range, our cognition apparently not being organized to bring the full set into attention.

One main limitation of corpus research, though, is that it in general can not directly yield many abstract linguistic patterns. One reason is that the sentences speakers produce largely are multiply elliptical, omitting constituents that could, in principle, fill in the complete potential structure of some construction or lexical form. The various occurrences of such a construction or lexical form in a corpus will thus typically each lack certain components of their potentially full structure, components whose values might be understood but are implicit (see Talmy 2000a: Ch. 4). For example, in one of its usages, a verb can have one or more optional arguments in its complement structure that are not overtly represented, and will have a particular aspectual signature that can lack some accompanying temporal expression to reveal it. Moreover, a lexical form commonly has several usages, so that a verb, for example, will have several different complement structures and aspects, all of them appearing scattered through a corpus without identification as to the particular usage in play. Unguided attempts to characterize a verb's semantic-syntactic pattern can readily become confounded without sorting out the different levels and scopes in effect. Introspection has the advantage here, since our cognition seems to be organized so as to abstract out the distinct full patterns associated with lexical forms and constructions.

Another limitation of a corpus is that it does not mark those occurrences of a lexical form or construction that are fully felicitous and well-formed, as against those that are less well selected or that deviate in any of the other respects indicated earlier – a recurrent feature of natural speech. This can lead to confounded characterizations of the properties of such forms and constructions. Again, our linguistic cognition is organized so as to have abstracted out the ideal grammatical and semantic properties of particular lexical forms and constructions, ones that can emerge through introspection in a deliberative process like writing, but that are commonly breached in the kind of fluent speech recorded in corpora.

5. Experimental method

As a methodology in the study of language, cognitive psychology is distinguished in its application of the experimental method to linguistic cognition. This method largely consists of presenting a number of individuals with stimuli or instructions, prepared with the aim of addressing a single cognitive factor, and monitoring their responses. Under this aegis, techniques range widely, from the use of instrumentation for special presentations of language-related stimuli or for recordings of physical responses to them; to instructions to generate a specified linguistic output, say, to produce all the words with a given meaning that one can think of in a brief period; to the re-presentation of certain linguistic stimuli after a lengthy period to test for memory. And the time scale of the cognitive processes such techniques probe ranges from the millisecond level to months, although perhaps the bulk of experiments aims at the shorter end.

One advantage of the experimental method is precisely this access to the millisecond scale of cognitive processes, which is not available to any other methodology. Toward the scale of whole seconds, the experimental method begins to share access with the methodology of audio- and videographic analysis.

Additional advantages of the experimental method can be regarded as complements to advantages found in other methodologies, where the experimental method in turn has limitations. Thus, if an advantage of audiovisual recordings and corpora is that they permit the examination of naturalistic speech, the complementary advantage that experimental psychology shares with first-condition introspection is that the researcher can carefully control the stimuli that evoke linguistic behavior. That is, he can probe the system of linguistic cognition, even perturb it, as a means for detecting aspects of its organization often obscured or sporadic as it functions naturalistically.

Further, if an advantage of introspection and of analytic thought is their direct access to subjectively presented objects of examination, with the researcher attending to the products of her own mind, the complementary advantage that experimental psychology shares with the use of live recordings and corpora is a focus on objectively presented objects of examination, with the researcher attending to the products of the minds of other individuals. The earlier-mentioned concern of some that the act of introspection might affect the object of attention is not resolved by the experimental method, which can be equally at risk in affecting the target of observation. But the concern that introspection might not be reliable due to its very subjectivity is addressed by the experimental method.

Finally, an advantage of audiovisual recording and introspection is that they both permit an in-depth examination of linguistic behavior within the cognition of a single individual, and are thus able to address cognition as an integrated system, one consisting of components in particular interactions. But then the complementary advantage that experimental psychology shares with the use of corpora is that they both base their conclusions on the linguistic behavior exhibited across a set of individuals, and are thus able to abstract away from individual differences and discern even slight characteristics that linguistic cognition tends toward in humans.

One main limitation of the experimental method in psychology might be found in this last contrast. The techniques designed to isolate what is taken as a single factor in linguistic cognition and to keep other factors among which it is embedded from confounding the probe can lead to decontextualization. The result can be an insufficient capacity to track the factor as it weaves through or interacts with its fellows within an integrated system, or to ascertain whether what has been isolated in fact constitutes a functionally discrete factor.

As can be seen, each of the methodologies now being applied to cognitive linguistics has unique capacities that make it necessary for our overall understanding of conceptual structuring in language, as well as having limitations that make the other methodologies additionally necessary for this understanding. It can be further observed that each methodology needs to look at the findings of its fellows for new ideas as to where to proceed next within its own practices. The chapters in this volume are both early contributions in this collegial spirit, and harbingers of collaborations to come.

References

Talmy, L. (2000a). *Toward a Cognitive Semantics, Volume I: Concept structuring systems.* Cambridge: MIT Press.

Talmy, L. (2000b). *Toward a Cognitive Semantics, Volume II: Typology and process in concept structuring.* Cambridge: MIT Press.

Talmy, L. (2003). The representation of spatial structure in spoken and signed language. In Karen Emmorey (Ed.), *Perspectives on Classifier Constructions in Sign Language* (pp. 169–195). Mahwah, NJ: Lawrence Erlbaum.

Talmy, L. (2004). Recombinance in the Evolution of Language. In Jonathon E. Cihlar, David Kaiser, Irene Kimbara, & Amy Franklin (Eds.), *Proceedings of the 39th Annual Meeting of the Chicago Linguistic Society: The Panels.* Chicago: Chicago Linguistic Society.

Talmy, L. (2006). The fundamental system of spatial schemas in language. In Beate Hamp (Ed.), *From perception to meaning: Image Schemas in Cognitive Linguistics* (pp. 199–234). Mouton de Gruyter.

Talmy, L. (Forthcoming). *The attentional system of language.* Cambridge: MIT Press.

Why cognitive linguists should care more about empirical methods

Raymond W. Gibbs, Jr.

1. Introduction

Linguistics and psychology have always had a curious relationship. Ever since the early days of generative linguistics when Chomsky started to argue that linguistics was a subfield of cognitive psychology, there has always been intense debate as to whether linguistic theories are "psychologically real." In the early and mid 1960s, for example, psychologists were quite enthusiastic about transformational grammar being part of the underlying principles organizing sentence processing. But a vast body of experimental research showed by the early 1970s that this was simply not the case (Fodor, Bever, & Garrett 1974). Since that time, psychologists have struggled to apply various linguistic theories to explain language acquisition, production, and comprehension, with many psychologists expressing significant skepticism toward any theory of language use that is not based on objective scientific experiments. This has most recently been true in regard to how psychologists view the various theories and claims of cognitive linguistics. Many psychologists suggest that linguistic intuitions alone, even those of trained linguists, are insufficient sources of evidence for establishing "what people ordinarily do" when using and understanding language (Glucksberg 2001; Murphy 1996; Veraeke & Kennedy 1996). The best, and in some people's view, the only, way to study ordinary language use is to objectively study the behavior of naïve human participants in controlled experimental settings.

My aim in this chapter is to present the case for why cognitive linguists should care more about empirical methods given the skepticism from people outside their field. First, I outline in a bit more detail some of the reasons for why the skilled intuitions of cognitive linguists may be useful, but not at all conclusive, in arguing for the specific influences of thought and embodied experience in everyday language use. Second, I suggest several principles that cognitive linguists should adopt in articulating psychologically plausible theories of mind and language. At the same time, I urge cognitive linguists to more fully explain the methods they use in analyzing linguistic phenomena and in making claims about human conceptual systems. I do not believe, contrary to some of my colleagues in psychology, that cognitive linguists must do experiments to have their ideas be considered as psychological theories. Nonetheless, there are various empirical, experimental techniques that are part of the arsenal of "indirect methods" used in psycholinguistics that have proven to be quite useful in providing support for many of cognitive linguists' claims about mind and language. I briefly outline several of these in the third part of this chapter. My overall goal is to provide ways of drawing cognitive linguists and psychologists closer together, while simultaneously respecting these scholars' different theoretical goals and empirical methods.

2. The problem with introspection

Despite their differences with generative linguists, cognitive linguists mostly employ traditional linguistic methods of examining native speakers' intuitions about the grammaticality and meaningfulness of linguistic expressions in order to uncover idealized speaker/hearer linguistic knowledge. In most cases, the linguistic expressions examined are made-up (i.e., not derived from actual spoken and written discourse), and the intuitions studied are those of the scholar actually conducting the work. Many linguists argue that their own intuitions about linguistic matters should count for something more than asking ordinary speakers who lack linguistic training. Within cognitive linguistics particularly, a scholar's trained intuitions seem essential in being able to uncover language-mind links, such as the mental spaces, the image schemas, the conceptual metaphors, and so on that have now become a major foundation for cognitive linguistic theories of human conceptual systems.

I personally have a split view about the kinds of practices that cognitive linguists engage in when doing their work. On the one hand, I continue to be impressed with the different systematic analyses of linguistic patterns that point to different underlying conceptual structures that may provide partial motivation for the existence of words, utterances, and discourse structures within contemporary language. Psychologists should not ignore these findings simply because they are not the products of experiments. Many of my own experimental studies within cognitive psychology and psycholinguistics suggest that cognitive linguistic conclusions about the nature of human conceptual systems may indeed be correct and thus psychologically real (Gibbs 1994, 2006; Gibbs, Lima, & Francuzo 2004). In this manner, the trained intuitions of cognitive linguists have provided detailed insights into possible language-mind-body interactions that serve as the source of experimental hypotheses on the workings of the cognitive unconscious.

Yet I share with my colleagues in Psychology, and other disciplines (see Sandra 1998; Sandra & Rice 1995), some skepticism about trusting cognitive linguists' arguments and conclusions because these are so heavily based on individual introspections about matters of linguistic structure and behavior. Although introspections can be valuable sources for constructing hypotheses, we must always be cautious in accepting any individual analyst's linguistic judgments. Linguists assume that each scholar's intuitions should be representative of all speakers of a language, because each person within a linguistic community presumably shares the same underlying linguistic competence (Psychology does this in psychophysics where only a few participants' perceptual judgments are presumably needed to establish the real workings of the visual system given the belief that everyone's visual system is alike). But there is considerable variation in linguists' introspections. For instance, different linguistic theories of idiomaticity often rest with scholars varying intuitions about the acceptability, and/or grammaticality, of different word strings (under different syntactic permutations). Not surprisingly, linguists' introspections on such matters often are most consistent with their own particular view of idiomaticity, and more generally, the interface between the grammar and the lexicon (see Bresnan, & Kaplan 1982; Gibbs 1994; Nunberg, Sag, & Wasow 1994). An outside observer may ask "Whose intuitions, and ultimately which theory, should I trust?"

The second concern with linguists' introspections has to do with the possibly biased nature of any one person's observations about the cognitive unconscious. Smart people like to believe that they can articulate the inner workings of their own minds. My undergraduate students in Psychology often report, after I have presented them some recent empirical findings about the nature of mind "Ray, my brain doesn't work like that!" as if they somehow have privileged access to their unconscious cognitive processes that we psychologists on the outside can never see. But psychological studies, across a wide range of subfields within the discipline, have long demonstrated that people actually have very poor insights into the underlying cognitive processes at work when they perceive, learn, solve-problems, use language, and, most interestingly, have different emotional reactions to their own predicaments and to other people (Wilson 2002). The fact that we think we can introspect about the inner workings of our minds does not mean that such intuitions, even if trained, are either consistent or accurate. Research from both social psychology and cognitive psychology shows that people often give explanations for their decisions which vary significantly from what is shown by more objective means (Wilson 2002), and that people can significantly vary from one day to the next in reporting their beliefs or knowledge, even for simple things like the names of all the birds or furniture they know (Barsalou 1993). People may sometimes have reasonable access to certain kinds of knowledge, such as some autobiographical events, but even here there are studies showing significant degrees of self-illusion about the accuracy of what one putatively knows with people often reporting as "it really happened" events that they only imagined (Thomas, Bulevich, & Loftus 2003).

Our conscious ideas about the workings of the unconscious mind may be flawed for a number of reasons, even for those individuals who are trained in providing detailed analyses of their intuitions, such as many linguists and philosophers. In general, the adaptive unconscious mind differs from the conscious mind along a number of different dimensions that have been understood through many years of scientific study (adapted from Wilson 2002):

Adaptive/cognitive unconscious	Consciousness
Multiple systems	Single system
Online pattern detector	After the fact check and balance
Concerned with the here and now	Taking the long view
Automatic, fast, unintentional	Slow, effortful, intentional
Uncontrollable	Controllled
Rigid	Flexible
Precocious	Slower to develop

This list of differences between the adaptive/cognitive unconscious reinforces the idea that is may be impossible to understand the operations of the unconscious mind through conscious introspection alone (i.e., a first-person approach). Even psychotherapy, which studies show can be quite effective, works more because it allows a person to construct a better conscious narrative about one's thoughts, feelings, and experiences than it does in providing deeper, and accurate, insights into unconscious mental functioning. One may argue that the unconscious and conscious minds are still part of the same overall system (i.e., the person) and therefore must work in some harmonious way together as part of some grand overall design. But even this idea may not necessarily be true, as many cognitive scientists now question whether consciousness has any direct bearing on unconscious mental processes (Libet 2004; Wegner 2002).

It is not surprising, then, that many cognitive scientists are skeptical of theoretical claims based simply on one's intuitions or introspections, no matter how well trained these may be. Cognitive psychologists, and others, criticize cognitive linguistic work because it is so heavily based on individual analysts' intuitions (i.e., cognitive linguists- a third-person approach), and thus does not constitute the kind of objective, replicable data preferred by many scholars in the cognitive and natural sciences (e.g., data collected on large numbers of naïve participants under controlled laboratory conditions). This desire for objective evidence, based on experiments that can be replicated, and that test falsifiable hypotheses (more on this below) is especially needed if one wishes to make generalizations about the way that people ordinarily, and automatically, engage in cognitive and linguistic processing. Cognitive psychologists argue that indirect methods (i.e., not based on first-person assessments of unconscious cognition) must be employed to examine what people do, and how they do it, without asking them to say what they are doing, precisely because we now know how unreliable such reports can be.

3. Do cognitive linguists use empirical methods?

Beyond the concern about the reliability of linguists' introspections, and whether it is possible to understand the cognitive unconscious mind through introspection, there is also the deeper problem of specifying exactly what it is that cognitive linguists do when they do their work. Consider a case close to my own research interests- identifying conceptual metaphors from the systematic analyses of linguistic expressions. For instance, read the following set of expressions.

> "Look how far we have come."
> "We are not making any progress with this research."
> "I am just spinning my wheels trying to get a Ph.D."
> "I am at a turning point in my life."

Since Lakoff and Johnson (1980), cognitive linguists have argued that these conventional expressions are not isolated, but are related in slightly different ways to a single underlying conceptual metaphor LIFE IS A JOURNEY. This conceptual metaphor is presumed to be part of people's ordinary conceptual system that functions automatically in how people conceive of themselves, and others', experiences. Linguistic research, across a wide-range of languages, including signed languages, now shows that conceptual metaphors are critical in motivating the creation and continued existence of systematic conventional expressions, polysemous words, many novel metaphors, and play a role in gesture (Gibbs 1994, 2006; Gibbs & Steen 1999; Lakoff & Johnson 1999).

Yet how accurate are these claims? Do ordinary speakers really have conceptual metaphors and use them automatically in everyday thought and language? How does one even establish that a given word or expression in context expresses metaphorical meaning? Part of the resistance to cognitive linguists' claims is that these scholars do not sufficiently explain the methods employed in doing their linguistic analyses, and most importantly in drawing inferences from systematic patterns of language (a problem by itself) to claims about the underlying nature of human conceptual systems. We may be impressed by possible relationships between so-called conventional expressions when these are presented out of context. But how does any scholar really determine what words and phrases express metaphorical meanings or reflect metaphorical concepts?

To get a better sense of these difficulties, consider the following short paragraphs from an editorial published in the San Francisco Chronicle, April 29, 2003 (A22), titled "Toward a new Iraq."

> The job of constructing a new, democratic Iraq from the social wreckage left by Saddam Hussein will take many months and a steely determination by U.S. sponsors of the process to stay focused on the rights of all Iraqis- and to maintain order in the country until those rights are sufficiently protected by a new government.
>
> In the meantime, improved security in the streets and the restoration of war-damaged services should help create a climate in which people can think about their political options beyond the task of just staying alive.
>
> President Bush sought to boost the democracy-building effort in a speech Monday to Iraqi Americans in Michigan. He walks a fine line in assuring that the United States has 'no intention of imposing our form of government or our culture,' but insisting that all Iraqis will enjoy a voice and legal protections.

What words and phrases in these excerpts are metaphorical? Some readers immediately point out that the word "Toward" in the editorial title is metaphorical in that the writer is not speaking of physically moving to a new place called Iraq, but is conceiving of metaphorically moving toward a new nation-state that emerges from the Iraq war. But what about the phrase "constructing a new Iraq"? Is this being used metaphorically, or might it simply refer to the physical rebuilding of Iraq after the devastation of the war and Hussein's long-time neglect of the country? Might this phrase have both a literal and metaphorical meaning? The term "social wreckage" seems metaphorical, or at least it does to some speakers. The adjective in the phrase "steely determination" seems quite metaphorical, precisely because "determination" is an abstract concept that has no physical dimensions. Finally, what about the preposition "on" in "stay focused on the rights of all Iraqis"? Is there something physical here that actually represents some contact between two entities, as in "The cat is on the mat"?

When asked, cognitive linguists will typically have strong responses to these important questions, and frequently explain, on a case-by-case basis, the reason for why, for example, a set of conventional expressions may be motivated by some underlying conceptual metaphor (or primary metaphor). Cognitive linguistics go on to argue that these methods are reliable, are taught regularly in linguistic classes, and have successfully illuminated many facets of language and mind that were undiscoverable by other linguistic methods. However, the remarkable fact is that there are very few published writings on methods in cognitive linguists (see Kövecses 2002 for an exception). For example, there is virtually no set of reliable, replicable methods that can be employed to identify words as metaphorical, or for relating systematic patterns of entire expressions to underlying conceptual metaphors. I am not claiming that cognitive linguists do not have empirical methods. But they really should place far more effort toward explicating their methods, and strive to show that the methods they employ are reliable, and replicable. On a personal note, the need for such explications is perhaps the single main complaint I encounter from metaphor scholars in many disciplines, ranging from applied linguistics to experimental psychology. Cognitive linguistics, as a discipline, would have much greater status within the cognitive sciences if they paid more attention to explicating the methods they use, and demonstrate that these provide for consistent, replicable research results.

4. Challenges for cognitive linguistics

In addition to trying to better explicate their methods for analyzing linguistic data, and better justifying their claims for different language-mind, and language-mind-body connections, cognitive linguists need to better frame their work so that it may be more amenable to experimental test. A common complaint from scholars outside of cognitive linguistics is that it is difficult to falsify aspects of theories within the discipline. Some cognitive linguists respond to these complaints by saying "That's not my problem or concern," while others go so far as to reject falsification as an important part of their theoretical work. Nonetheless, cognitive linguists still strongly maintain that their research provides detailed accounts of linguistic and cognitive behavior, and as such should have scientific credibility. Even if cognitive linguists do not conduct experiments, their work would significantly benefit from adherence to several general principles in framing their theories and research implications (Gibbs 2000).

First, different hypotheses must be falsifiable! Thus, each hypothesis must be stated in such a way that it can be experimentally/empirically examined and shown to be possibly false (and if not shown to be false, then one can reject the null hypothesis and conclude that there is evidence in support of the hypothesis). The problem of falsifying theories/ideas from cognitive linguistics is a big problem, and leads me to remain somewhat skeptical about certain claims (e.g., from conceptual blending theory). Ideas are very appealing, but it is unclear how one would go out and test this as compared to reasonable alternative hypotheses.

This point leads to the second recommendation- consider alternative explanations. For instance, might there be alternative reasons for the apparent systematicity among conventional expressions? Might systematicity just be a historical product, but have no role at all in how contemporary speakers think and use language? Might the systematicity among various words and expressions be a matter of polysemy, instead of conceptual metaphor, as some psychologists have claimed, incorrectly in my view (Glucksberg 2001; Murphy 1996). An example of the failure to consider alternative hypotheses in cognitive linguistics is seen in some, but not all, work on conceptual blending theory (Fauconnier & Turner 2002). Conceptual blending theory predicts that various sorts of blending processes should occur when people understand certain kinds of complex linguistic expressions (Coulson 2001). One can go out and do an experiment which shows that, indeed, people take longer to process certain utterances compared to others, or that some parts of utterances, where blending should occur, specifically take extra time to comprehend or engage more complex brain activity. But many other theories of linguistic processing would predict the very same finding! Thus, it is not clear that conceptual blending theory, despite its different conceptual and terminological perspective, is sufficiently unique to be considered the most viable psychological theory. Making the case for the "psychological reality" of any cognitive linguistic theory demands that such arguments be situated within the context of ongoing debates, and alternative theories within cognitive science.

Finally, cognitive linguists must realize that language understanding is not a single kind of mental process. Thus, the kind of mental activity used when a person listens to real speech, or reads a text in real-time, is quite different from the processes involved when a person reflects on what one is hearing or reading. This too is a major concern and perhaps the main reason why many cognitive scientists, especially in psychology, are deeply skeptical of ideas from cognitive linguistics. For example, cognitive linguists have written that conceptual metaphors are "used constantly and automatically, with neither effort or awareness" (Lakoff 1993). But is this true? Does the linguistic evidence alone provide the right kind of evidence to judge this idea? Many say no (see Glucksberg 2001; Gibbs 1994).

What is needed, then, is a more detailed set of specific hypotheses that can be individually examined using, perhaps, different experimental techniques. Among the possible hypotheses are (see Gibbs 1994; Katz, Cacciari, Gibbs, & Turner 1999):

1. Conceptual metaphors motivate why certain words and expressions have acquired their various figurative/metaphorical meanings over time (i.e., diachronically), but play no role in how contemporary speakers use and understand conventional and novel metaphorical expressions.
2. Conceptual metaphors motivate why certain words and expressions have their specific figurative meanings within linguistic communities and contemporary speakers can under the right circumstances, determine these motivations. Thus, knowledge of conceptual metaphors reflects something about idealized speakers-hearers. BUT conceptual metaphors are not "psychologically real" in the sense of being parts of ordinary, contemporary speakers' conceptual systems.
3. Conceptual metaphors motivate why certain words and expressions have their specific figurative meanings and these metaphors underlie why contemporary speakers tacitly recognize why these words and phrases have the particular meanings they do. Thus, conceptual metaphors are part of ordinary speakers' conceptual systems. But

conceptual metaphors are not necessarily employed "automatically" each and every time people use and understand particular kinds of language.

4. Conceptual metaphors motivate why certain words and expressions have the meanings they do, are part of speakers' conceptual systems and enable people to recognize something of why these words and phrases have the meanings they do AND are employed automatically each and every time when people use and understand language.

These different hypotheses must be examined by appropriate empirical methods. Thus, 1 and 2 are surely within the domain of cognitive linguistics research. But 3 and 4 require the "indirect methods" of cognitive psychology/psycholinguistics. These methods are, again, "indirect" in that they do not require people to introspect about their own, mostly unconscious, mental processes. Rather, the right method will provide data that enables the researchers to draw inferences about underlying mental processes (e.g., people automatically accessing tacit conceptual metaphors during on-line metaphor comprehension). My point here, more generally, is that cognitive linguists must be sensitive to the different levels at which "linguistic understanding" can be studied and explained, and recognize that their own methods of systematic, conscious analysis of linguistic expressions cannot provide the needed insights into "automatic" language production or processing.

5. Examples of relevant methods

Let me now briefly describe some methods that experimental psycholinguists have successfully employed in testing various implications of cognitive linguistic ideas, primarily about conceptual metaphors, as described above. These various techniques are aimed at examining hypotheses 3 and 4 above.

5.1 Mental imagery

The first method for examining hypothesis 3 is to investigate people's mental imagery for conventional phrases. For instance, do people know why the expression "spill the beans" has the figurative meaning, "reveal the secret." People are poor at answering this question, but one can elicit people's mostly unconscious knowledge about, in this case, conceptual metaphors, using a more indirect method by having people form mental images for linguistic expressions (Gibbs & O'Brien 1990; Gibbs, Strom, & Spivey-Knowlton 1997). Consider the idiom "spill the beans." Try to form a mental image for this phrase and then ask yourself the following questions. Where are the beans before they are spilled? How big is the container? Are the beans cooked or uncooked? Is the spilling accidental or intentional? Where are the beans once they've been spilled? Are the beans in a nice, neat pile? Where are the beans supposed to be? After the beans are spilled, are they easy to retrieve?

Most people have definite responses to these questions about their mental images for idioms. They generally say that the beans were in some pot that is about the size of a person's head, the beans are uncooked, the spilling of the beans is accidental, the spilled beans are all over a floor and are difficult to retrieve. This consistency in people's intuitions about their mental images is quite puzzling if one assumes that the meanings of idioms are arbitrarily determined. People's descriptions about their mental images for idioms reveal some of the metaphorical knowledge that motivates the meanings of idiomatic phrases. One study examined people's mental images for groups of idioms with similar figurative meanings, such as anger (e.g., "blow your stack," "hit the ceiling," "flip your lid") (Gibbs & O'Brien 1990). Participants were asked to describe their mental images for these idioms and to answer questions about the causes, intentionality, and manner of actions in their mental images for these phrases.

Not surprisingly, people give many different responses across the different idioms presented, and one challenge for researchers is to systematically categorize these into different, meaningful groups. Psychologists are reasonably good at coding different human behaviors, but experience greater difficulty analyzing naturalistic linguistic expressions. This is one place where my own study of cognitive linguistics has served me quite well in helping me to do experimental research.

Gibbs and O'Brien (1990) actually found that participants' descriptions of their mental images were remarkably consistent for different idioms with similar figurative meanings. The general schemas underlying people's images were not simply representative of the idioms' figurative meanings, but captured more specific aspects of the kinesthetic events with the images. For example, the anger idioms such as "flip your lid" and "hit the ceiling" all refer to the concept of "getting angry," but participants specifically imagined for these phrases some force causing a container to release pressure in a violent manner. There is nothing in the surface forms of these different idioms to tightly constrain the images participants reported. After all, lids can be flipped and ceilings can be hit in a wide variety of ways, caused by many different circumstances. But the participants' protocols in this study revealed little variation in the general events that took place in their images for idioms with similar meanings.

Participants' responses to the questions about the causes and consequences of the actions described in their images were also highly consistent. Consider the most frequent responses to the probe questions for the anger idioms (e.g., "blow your stack," "flip your lid," "hit the ceiling"). When imagining anger idioms, people reported that pressure (i.e., stress or frustration) causes the action, that one has little control over the pressure once it builds, its violent release is done unintentionally (e.g., the blowing of the stack) and that once the release has taken place (i.e., once the ceiling has been hit, the lid flipped, the stack blown), it is difficult to reverse the action. We speculated that people's images for the anger idioms are based on folk conceptions of certain physical events. That is, people use their embodied knowledge about the behavior of heated fluid in containers (e.g., the bodies as containers and bodily fluids within them) and map this knowledge onto the target domain of anger to help them conceptualize in more concrete terms what is understood about the concept of anger. Various specific entailments result from these general metaphorical mappings, ones that provide specific insight into people's consistent responses about the causes, intentionality, manner, and consequences of the activities described by stacks blowing, lids flipping, ceilings being hit and so on.

We did not claim that people necessarily form mental images during ordinary idiom comprehension. But asking people to form mental images, and answer specific questions about them, reveals significant constraints that conceptual metaphors play in motivating why conventional phrases have the meanings they do. Thus, conceptual metaphors appear to be the main link between many idioms and their figurative meanings. Once more, this tacit knowledge could not be uncovered by simply asking people about why idioms mean what they do. Yet the indirect method of forming mental images can provide such insights.

5.2 Context-sensitive judgments about metaphorical meaning

A different method for examining hypothesis 3 is to assess people's judgments of similarity between idioms and different discourse contexts. Nayak and Gibbs (1990) hypothesized that contexts provide information about specific metaphoric mappings that cue readers to the specific figurative meanings of idioms. Participants in one experiment were presented with short scenarios about a particular emotion concept that were constructed to prime one of the metaphorical mappings inherent in its prototypical structure. Consider the following example:

> Mary was very tense about this evening's dinner party. The fact that Bob had not come home to help was making her fume. She was getting hotter with every passing minute. Dinner would not be ready before the guests arrived. As it got closer to five o'clock the pressure was really building up. Mary's tolerance was reaching its limits. When Bob strolled at ten minutes to five whistling and smiling, Mary
>
> a. blew her stack
> b. bit his head off

The story was written to prime the metaphorical mapping ANGER IS HEAT IN A PRESSURIZED CONTAINER by depicting Mary's increasing anger in terms of increasing pressure and heat. The use of phrases such as "very tense, making her fume, getting hotter, the pressure was really building up" and "reaching it's limits" are specific references to this mapping. Participants rated the appropriateness of each idiom ending for the given scenario. If people access the metaphoric mapping reflected in an idiom's lexical structure, they should interpret "blew her top" as being more appropriate than "bit his head off" even though both phrases are grammatically and conceptually (at the same stage of the prototype) appropriate for the given scenario.

But now consider a slightly different scenario that primes a different conceptual metaphor, ANGRY BEHAVIOR IS ANIMAL BEHAVIOR, and should result in different expectations:

> Mary was getting very grouchy about this evening's dinner party.
> She prowled around the house waiting for Bob to come home to help.
> She was growling under her breath about Bob's lateness.
> Her mood was becoming more savage with every passing minute.
> As it got closer to five o'clock Mary was ferociously angry with Bob.
> When Rob strolled in at 4;30 whistling and smiling, Mary
> bit his head off
> blew her top

In this case, "bit his head off" appears to be more appropriate than in the earlier contexts because the mental model is structured according to the metaphor ANGRY BEHAVIOR IS ANIMAL BEHAVIOR. This suggests that idioms must reflect the same metaphorical mapping information as its context to be considered most appropriate. In fact, the results clearly showed that the metaphoric mappings underlying idiomatic phrases affect participants' interpretation of the meanings and appropriate use of these figurative expressions. Participants were sensitive to the congruence between the metaphoric information in idioms and contexts. It appears that the mapping of the conceptual information in discourse contexts to people's knowledge about conceptual metaphors determines readers' intuitions about the appropriate use of idioms. These findings provide experimental evidence in support of hypothesis 3 that conceptual metaphors influence people's interpretation of why idioms mean what they do and are used in specific discourse contexts.

5.3 Embodied intuitions and metaphorical inferences

One of the reasons why cognitive psychologists are skeptical of cognitive linguistic work is because of the inherent circularity in reasoning from language to underlying concepts to language again. Cognitive psychologists seek ways of stepping outside of the language to language circle by having independent ways of predicting in advance something about linguistic meaning, as opposed to postulating backward-looking reasons or motivations for why some specific word or phrase has the meaning it does. One strategy for doing this in respect to hypothesis 3 is to look independently at people's nonlinguistic knowledge about source domains and then use this to make predictions about the meanings of metaphorical phrases referring to target domains. My experimental strategy to see if this might be true was to make specific predictions about what various idioms, say those motivated by ANGER IS HEATED FLUID IN A CONTAINER, actually mean by looking at the inferences that arise from the mapping of people's nonlinguistic knowledge of heated fluid in a container onto the idea of anger (Gibbs 1992).

Participants in this study were asked about their understanding of events corresponding to particular source domains in various conceptual metaphors (e.g., the source domain of heated fluid in a container for ANGER IS HEATED FLUID IN A CONTAINER). For instance, participants were asked to imagine the embodied experience of a sealed container filled with fluid, and then they were asked something about causation (e.g., "What would cause the container to explode?"), intentionality (e.g., "Does the container explode on purpose or does it explode through no volition of its own?"), and manner (e.g., "Does the explosion of the container occur in a gentle or a violent manner?").

Participants gave highly consistent responses to these questions. Thus, people responded that the cause of a sealed container exploding its contents out is the internal pressure caused by the increase in the heat of the fluid inside the container, that this explosion is unintentional because containers and fluid have no intentional agency, and that the explosion occurs in a violent manner. This provides a rough, nonlinguistic profile of people's understanding of a particular source domain concept (i.e., "image-schematic structures") of the source domains.

If hypothesis 3 is correct, people's intuitions about various source domains should then map onto their conceptualizations of different target domains in very predictable ways. Not surprisingly, when people understand anger idioms, such as "blow your stack," "flip your lid," or "hit the ceiling," they inferred that the cause of anger is internal pressure, that the expression of anger is unintentional, and is done in an abrupt violent manner. People did not draw the same inferences about causation, intentionality, and manner when comprehending literal paraphrases of idioms, such as "get very angry." Additional experiments showed that people find idioms to be more appropriate and easier to understand when they are seen in discourse contexts that are consistent with the various entailments of these phrases, which, again, were predicted in advance from the nonlinguistic analysis of the source domain concepts. In general, these psycholinguistic studies are significant for hypothesis 3 because they provide independent, nonlinguistic ways of predicting something about the specific metaphorical meanings some linguistic expressions are likely to possess. These psychological findings are hard to reconcile with the view that the figurative meanings of idioms are determined only on the basis of their individual lexical items or have the meanings they do for arbitrary, or historically opaque reasons. Contemporary speakers appear to have tacit intuitions about their metaphorical understanding of certain abstract concepts that lead them to talk about these concepts in particular metaphoric ways. No other theory of idiomaticity comes close to being able to describe exactly why it is that idioms have the very specific meanings they do for contemporary speakers or why people appear to quickly draw specific inferences about what idioms mean.

5.4 Not all methods work!

In all fairness, the debate over conceptual metaphors in cognitive psychology has provided evidence that seems contrary to some of the putative predictions of cognitive linguistics. Consider the work of McGlone (1996) who examined people's verbal paraphrases for linguistic metaphors. Participants in a first experiment paraphrased verbal metaphors, such as "The lecture was a three-course meal." Only 24% of these paraphrases contained any references consistent with underlying conceptual metaphors, such as IDEAS ARE FOOD. Even when participants were asked to give figurative paraphrases of the verbal metaphors, they still most frequently produced paraphrases inconsistent with related conceptual metaphors. Thus, when given the verbal metaphor "Dr. Moreland's lecture was a three-course meal for the mind," only 1/3 of the paraphrased mentioned source domain terms (e.g., food) related to the conceptual metaphor IDEAS ARE FOOD. Nonetheless, almost all of the metaphorical paraphrases reflected some recognition of the stereotypical properties of three-course meals that might be attributed to lectures, such as "large quantity," and "variety." A third study asked participants to rate the similarity between different metaphorical expressions. The data showed that people do not perceive expressions motivated by conceptual metaphor to be any more similar in meaning than they did expressions motivated by different conceptual metaphors. Thus, "Dr. Moreland's lecture was steak for the mind" was not seen as more similar to "Dr. Moreland's lecture was a three-course meal for the mind" than was "Dr. Moreland's lecture was a full tank of gas for the mind." A final study showed that conceptual metaphors consistent with a verbal metaphor were not better recall cues for participants trying to remember the verbal metaphors than were unrelated cues. Overall, the findings from these studies were taken to imply that people's interpretations of verbal metaphors are not necessarily related to their putative, underlying conceptual metaphors.

McGlone's data are interesting in many respects, although they are not especially surprising. First, it is not clear that having people verbally paraphrase a metaphor is the best method for tapping into different types of, possibly metaphorical, knowledge that might be used when people interpret, or make sense of, verbal metaphors. After all, various others empirical methods have shown some influence of conceptual metaphors on comprehension of, at least, idiomatic and proverbial phrases. One shouldn't imply that the failure to find effects using one task invalidates the positive evidence in favor of hypothesis 3 using different tasks unless some principled reasons are given for preferring one task over another. Paraphrase tasks are notoriously insensitive as measures of people's, especially children, ability to understand metaphors.

6. Bodily movement and metaphor comprehension

I now turn to two instances of methods for exploring the plausibility of hypothesis 4, namely that conceptual metaphors influence people's immediate comprehension of conventional, metaphorical phrases. Imagine that one hears the idiomatic expression "John blew his stack" in a conversation in which it is clear that the speaker's intended meaning is roughly "John got very angry." The figurative meaning of "blew his stack" is partly motivated by the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER. The question is whether people compute or access some conceptual representation for ANGER IS HEATED FLUID IN A CONTAINER when they immediately process the figurative meaning of "John blew his stack." Participants in one series of studies read stories one line at a time on a computer screen. Each story ended with an idiom ("John blew his stack"), a literal paraphrase of the idiom ("John got very angry"), or an unrelated literal statement ("John saw the dented door") (Gibbs, Bogdonovich, Sykes & Barr 1997). The computer measured how long it took people to read each line and then push a button signifying that they had understood what they just read. After reading the last line, participants were presented with a letter string and asked to decide as quickly as possible if this was a meaningful word in English. These letter-strings reflected either something about the conceptual metaphors underlying these idioms (e.g., "heat" for ANGER IS HEATED FLUID IN A CONTAINER having just read "John blew his stack") or letter-strings that were unrelated to these conceptual metaphors (e.g., "lead").

There were two important findings. First, people were faster to make lexical decision responses to the related metaphor targets (i.e., "heat) having just read idioms than they were to either literal paraphrases of idioms (e.g., "John got very angry") or control phrases (e.g., phrases still appropriate to the context such as "John saw many dents"). Second, people were faster in recognizing related metaphorical targets than unrelated ones having read idioms, but not literal paraphrases or unrelated phrases. This pattern of results suggests that people are immediately computing or accessing at least something related to the the conceptual metaphor ANGER IS HEATED FLUID IN A CONTAINER when they read idioms.

In another experiment, participants were faster to make lexical decision responses to metaphor targets (e.g., "heat") having read an idiom motivated by a similar conceptual metaphor (e.g., "John blew his stack") than an idiom with roughly the same figurative meaning but motivated by a different conceptual metaphor (e.g., "John bit her head off" which is motivated by the conceptual metaphor ANGER IS ANIMAL BEHAVIOR). People were also faster to respond to related targets having read idioms motivated by similar conceptual metaphors than when they read idioms motivated by different conceptual metaphors. In general, these online priming studies reveal that people appear to compute or access the relevant conceptual metaphor for an idiom during some aspect of their immediate processing of these phrases. It is not clear from these results whether the activated conceptual metaphor is used to interpret an idiom's meaning, or whether conceptual metaphors are simply tagged onto different idioms without serving as the causal basis for interpreting these conventional phrases. Nonetheless, this kind of data, and the methods involved in collecting it, is exactly what is required to test hypothesis 4.

A different, more recent, line of research investigated the possible influence of bodily action on people's speeded processing of simple metaphoric phrases, as "stamp out a feeling," "push an issue," "sniff out the truth" and "cough up a secret," each of which denote physical actions upon abstract items. Wilson and Gibbs (2006) hypothesized that if abstract concepts are indeed understood as items that can be acted upon by the body, then performing a related action should facilitate sensibility judgments for a figurative phrase that mentions this action. For example, if participants first move their arms and hands as if to grasp something, and then read "grasp the concept," they should verify that this phrase is meaningful faster than when they first performed an unrelated body action. Our hypothesis was that engaging in body movements associated with these phrases should enhance the simulations that people create to form a metaphorical understanding of abstract notions, such as "concept," even if "concepts" are not things that people can physically grasp. People's conceptual understandings of what a "concept" is, for example, need not be completely embodied and metaphorical. However, our suggestion is that some simulated construals of "concept" are rooted in embodied metaphor that may be highlighted by engaging in body actions relevant to what people mentally do with ideas.

Participants in this study first learned to perform various specific bodily actions (e.g., throw, stamp, push. swallow, cough, grasp) given different nonlinguistic cues. Following this, participants were individually seated in front of a computer screen. The experiment consisted of a series of trials where an icon flashed on the screen, prompting the participant to perform the appropriate bodily action. After doing this, a string of words appeared on the screen and participants had to judge as quickly as possible whether that word string was "sensible."

Analysis of the speeded sensibility judgments showed that participants responded more quickly to the metaphorical phrases that matched the preceding action (e.g., the motor action grasp was followed by "grasp the concept"), than to the phrases that did not match the earlier movement (e. g, the motor action kick was followed by "grasp the concept"). People were also faster in responding to the metaphor phrases having performed a relevant body moment than when they did not move at all. In short, performing an action facilitates understanding of a figurative phrase containing that action word, just as it does for literal phrases. A second study showed that same pattern of bodily priming effects when participants were asked to imagine performing the actions before they made their speeded responses to word strings. This result reveals that real movement is not required to facilitate metaphor comprehension, only that people mentally simulate such action.

Most generally, people do not understand the nonliteral meanings of these figurative phrases as a matter of convention. Instead, people actually understand "toss out a plan," for instance, in terms of physically tossing something (i.e., plan is viewed as a physical object). In this way, processing metaphoric meaning is not just a cognitive act, but involves some imaginative understanding of the body's role in structuring abstract concepts. People may create embodied simulations of speakers' messages that involve moment-by-moment "what must it be like" processes that make use of ongoing tactile-kinesthetic experiences. These simulation processes operate even when people encounter language that is abstract, or refers to actions that are physically impossible to perform.

7. Conclusion: Cognitive linguists need not do experiments

Cognitive linguistics is firmly embedded within the cognitive sciences, and as such is both a disciplinary and interdisciplinary endeavor. The interdisciplinary side of cognitive linguistics is evident in the increasing body of research in which linguists have collaborated with scholars from other disciplines, or have started to engage in research utilizing experimental and computational methods. I now talk with many younger cognitive linguistics students who are quite interested in doing informal experiments to test their ideas as part of their dissertation projects, in some cases using some of the methods described above, such as mental imagery and context-matching tasks. This is obviously a good thing for the field of cognitive linguistics overall, and for our understanding of human thought and language more generally.

However, my personal belief is that cognitive linguists need not become experimental psychologists or computer scientists for their work and ideas to be seen as legitimate with significant theoretical implications. There is a trend in cognitive science in which scholars in any one discipline always turn toward the right to seek evidence from a neighboring field to find additional, usually more empirical, support for their ideas and theories. For instance, philosophers often turn to linguistics, linguistics has historically turned to developmental and cognitive psychology, linguistics and psychology have often turned toward computer science, and most recently, cognitive scientists of all colors have turned toward neuroscience. Once more, these developments are natural and in many cases lead to important new work and empirical findings. But cognitive linguists are skilled at being able to conduct these sorts of systematic analyses, even if their methods for doing this are not always explicit, and have provided a huge body of work that simply could not be done by people in any other field. Why ask cognitive linguists to turn away from what they do best to secure their work on a different empirical foundation? My research has benefited greatly from cognitive linguistics studies, and we need more of this work and would hate to see cognitive linguists all try to become experimental psychologists, computer scientists, or neuroscientists. Doing experiments is hard work, and one does not casually pick up the skills needed to engage in this kind of research. What is needed, again, is for cognitive linguists to be more sensitive to some of the important properties of framing experimental hypotheses (e.g., constructing falsifiable hypotheses, considering alternative hypotheses), and trying to articulate their ideas, and empirical findings in ways that may be tested by scholars in other disciplines. This does not mean, however, that cognitive linguists must themselves run out and be something that they are not.

Finally, I have focused in this chapter on why cognitive linguists should care more about empirical methods, and suggested some of the ways that they could alter their work to better situate their findings within cognitive science. Yet psychologists, at the same time, would greatly benefit from learning more about cognitive linguistics, and learning to conduct some of the systematic analysis of linguistic expressions that is critical to understanding the conceptual/embodied motivation for linguistic meaning. Doing cognitive linguistics is, of course, hard work also. But the best way to appreciate the insights from cognitive linguistics, and apply these ideas to experimental tests, is to do cognitive linguistics. Some of us need help in doing such work, and my hope is that cognitive linguists will put more effort into sharing their knowledge and working methods with scholars from other disciplines.

References

Barsalou, L. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A. Collins & S. Gathercole (Eds.), *Theories of memory* (pp. 29–101). Hillsdale, NJ: Erlbaum.

Bresnan, J. & Kaplan, R. (1982). Grammars as mental representations of language. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. xvii–liii). Cambridge, MA: MIT Press.

Coulson, S. (2001). *Semantic leaps*. New York: Cambridge University Press.

Fauconnier, G. & Turner, M. (2002). *The way we think*: *Conceptual blending and the mind's hidden complexities.* New York: Basic Books.

Fodor, J., Bever, T., & Garrett, M. (1974). *Psycholinguistics*. New York: Wiley.

Gibbs, R. (1992). What do idioms really mean? *Journal of Memory and Language, 31*, 485–506.

Gibbs, R. (1994). *The poetics of mind: Figurative thought, language, and understanding*. New York: Cambridge University Press.

Gibbs, R. (2000). Making good psychology out of blending theory. *Cognitive Linguistics, 11*, 347–358.

Gibbs, R. (2006). *Embodiment and cognitive science*. New York: Cambridge University Press.

Gibbs, R. & O'Brien, J. (1990). Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition, 36*, 35–64.

Gibbs, R. & Steen, G. (Eds.). (1999). *Metaphor in cognitive linguistics*. Amsterdam: Benjamins.

Gibbs, R., Bogdonovich, J., Sykes, J., & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language, 37*, 141–154.

Gibbs, R., Lima, P., & Francuzo, E. (2004). Metaphor is grounded in embodied experience. *Journal of Pragmatics*, *36*, 1189–1210.

Gibbs, R., Strom, L., & Spivey-Knowlton, M. (1997). Conceptual metaphor in mental imagery for proverbs. *Journal of Mental Imagery, 21*, 83–110.

Glucksberg, S. (2001). *Understanding figurative language*. New York: Oxford University Press.

Katz, A., Cacciari, C., Gibbs, R., & Turner, M. (1998). *Figurative language and thought*. New York: Oxford University Press.

Kövecses, Z. (2002). *Metaphor: A practical introduction*. New York: Oxford University Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought-, Vol. 2* (pp. 201–233). New York: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. Chicago: University of Chicago Press.

Libet, B. (2004). *Mind time*. Cambridge, MA: Harvard University Press.

McGlone, M. (1996). Conceptual metaphor and figurative language understanding: Food for thought? *Journal of Memory and Language, 35*, 544–565.

Murphy, G. (1996). On metaphoric representations. *Cognition, 60*, 173–204.

Nayak, N. & Gibbs, R. (1990). Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General, 119*, 315–330.

Nunberg, G., Sag, I., & Wasow, T. (1994). Idioms. *Language, 70*, 491–538.

Sandra, D. (1998). What linguists can and can't tell about the human mind: A reply to Croft. *Cognitive Linguistics, 9*, 361–378.

Sandra, D. & Rice, S. (1995). Network analyses of prepositional meaning: Mirroring the mind- the linguist's or the language user's? *Cognitive Linguistics, 6*, 89–130.

Thomas, A., Bulevich, J., & Loftus, E. (2003). Exploring role of repetition and sensory elaboration in the imagination inflation effect. *Memory & Cognition, 31*, 630–640.

Vervaeke, J. & Kennedy, J. (1996). Metaphors in thought and language: Falsification and multiple meanings. *Metaphor and Symbol, 11*, 273–284.

Wegner, D. (2002). *The illusion of conscious will*. Cambridge, MA: Harvard University Press.

Wilson, N. & Gibbs, R. (2006). Real and imagined body movement primes metaphor comprehension. Manuscript submitted for publication.

Wilson, T. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.

They actually said that?

An introduction to working with usage data through discourse and corpus analysis

Irene Mittelberg, Thomas A. Farmer and Linda R. Waugh

1. Introduction

Language comes in all shapes and sizes, and as such, the methods employed by the language scientist must be able to accommodate this variability. The types of data that can illuminate the relationship between language and cognition fall anywhere on a broad spectrum anchored at the one end by data derived through rigorous experimental methodology and on the other end by data stemming from authentic language use. Here, we focus on the merits of discourse and corpus analysis within cognitive/functionalist linguistics, and will provide an overview of some of the many ways to analyze, both qualitatively and quantitatively, several types of usage-based data. Given constraints on space, we cannot cover the large, and arguably intimidating, body of work on discourse and corpus analysis. Instead, our goal is to highlight some of the central themes and contributions of each of these interrelated approaches, and to provide the readers with a wealth of resources to guide them in future endeavors. It should be noted that the points of view expressed here are influenced by established theoretical and empirical advances in the interplay between language and cognition (e.g., Barlow & Kemmer 2000; Bybee & Hopper 2001; Chafe 1994; Clark 1996; Gibbs 1994; Lakoff & Johnson 1999; Langacker 2001; Lantolf & Thorne 2006; Slobin 1996; Sweetser 1990; Tomasello 1998, 2003).

Different methodologies will be discussed in light of their potential to yield further empirical evidence that may substantiate theories currently deployed and debated within Cognitive Linguistics. We will explore the ways in which the systematic study of natural language usage can provide insights not only into the nature and specific organization of the linguistic system, but also into the interplay between linguistic, cognitive, social, and cultural phenomena. We will present some selected research strategies that can help the reader deal with methodological challenges. As will become evident, there are different tendencies to combine qualitative and quantitative methodologies. As is the case with other chapters in this current volume, the approaches discussed below are not all situated within Cognitive Linguistics, per se, but may provide ways to investigate cognitive and linguistic structure grounded in situated language use. Regardless of the method used, it will always be important to follow the appropriate protocols in the use of human subjects for data collection (see Gonzalez-Marquez et al. and Wilcox & Morford both, this volume.)

What one understands discourse to be depends very much on one's philosophy of language and one's goals in language study. Given the pervasiveness of discourse as mediator of, among other things, knowledge, social relations, and ideology, discourse analysis is a vast field that subsumes many different traditions and disciplines with their respective questions and methodologies (comprehensive overviews are provided in Beaugarde 1997; Jaworski & Coupland 1999; Johnstone 2002; Renkema 2004; Schiffrin 1994; Schiffrin, Tannen, & Hamilton 2001; Steen 2004a; van Dijk 1997a, b; Wetherell, Taylor, & Yates 2001). Accordingly, discourse can be viewed as "language above the sentence or above the clause" (Stubbs 1983: 1), the study of utterances (Schiffrin 1994: 39–41), "language in use" (Fasold 1990: 65), "a type of social practice" (Fairclough 1992: 28), as well as a process that continuously and recurringly shapes, and is shaped by, the ecological embeddedness of human experience (cf. Jaworski & Coupland 1999; Kramsch 2002; Schiffrin 1994). From a cognitive perspective, discourse, and its tight relation with cognitive processing and conceptual structure, has been conceived of as the socially constructed lieu of usage events from which linguistic units are abstracted, "retaining as part of their value any recurring facet of the interactive and discourse context. Linguistic structures thus incorporate discourse expectations and are interpretable as instructions to modify the current discourse state" (Langacker 2001: 143; see also Barlow & Kemmer 2000; Cienki, Luka, & Smith 2001; Koenig 1998; Langacker 1987; Liebert, Redeker, & Waugh 1997; Tomasello 1998, 2003; van Hoek, Kibrik, & Noordman 1999). In each case, the methodologies and technologies the researcher may employ will depend not only on the modality of the discourse data (spoken, written, signed, co-speech gesture, etc.), but also on the genre (conversation, interview, newspaper texts, academic prose, etc.). Some of these data types and genres are covered in the present volume: Grondelaers et al. worked with a corpus of written texts in both Netherlandic and Belgian Dutch, Waugh et al. with French and American English conversational data. Furthermore, Sweetser analyzed American English conversational and expository co-speech gesture data, and Mittelberg worked with multimodal data collected in linguistics courses.

A corpus can be viewed most simply as a database of concrete linguistic utterances, be they spoken, written, gestured, or signed. As noted by McEnery and Wilson (1996), desirable features of a corpus include machine-readability and a maximally representative sample of the targeted language population. Most commonly, corpora are used quantitatively to extract frequency information, but they are also used to identify lexical and syntactic patterns occurring in a language sample. Corpora are used in many subdisciplines of the language sciences as a source of usage data with potential ramifications for theoretical and applied linguistics (Aijmer & Stenström 2004; Biber, Conrad, & Reppen 1998; Coffin, Hewings, & O'Halloran 2004; Grondelaers et al. this volume; Kennedy 1998; Leistyna, & Meyer 2003; Meyer 2002; Sinclair 1991; Stubbs 1996; Waugh et al. this volume; Wilson, Rayson, & McEnery 2003). Below, we present a list of ways in which corpus-based research may be conducted within various sub-fields of the language sciences (taken in part from Barodal 2005):

– Lexicography: word meanings, frequency, distribution, collocation patterns.
– Grammar: use, distribution, and function of specific constructions across registers, synonymy, etc., reference grammars.
– Lexico-grammar: relation between lexical items and grammatical items/constructions.
– Language variation/registers: variation of linguistic features within and across registers; similarities and differences between spoken and written registers and modalities (Biber 2003).
– Language acquisition and development: L1 & L2 acquisition of specific linguistic features or patterns across learners (CHILDES, learner corpora, parallel corpora).
– Historical linguistics: changes in language structure/vocabulary/use across historical periods, register development, influence of sociolinguistic variables, such as gender, on language use, etc.
– Natural language processing/computational linguistics: tagging, parsing, information retrieval, etc. (Wallington, Barnden, Buchlovsky, Fellows, & Glasby 2004).
– Metaphor research: systematic/genre-typical metaphors (Cameron & Low 1999; Deignan 1999; Partington 1998; Steen 1994; Stefanowitsch & Gries 2006).
– Applied corpus linguistics: Dictionaries and grammar, the study of ideology and culture, translation, stylistics, forensic linguistics, language teaching (data-driven learning, parallel concordances, synonymy, phrase patterns, teaching methodologies, learner corpora), and literature (Hunston 2002; Partington 1998; Steen 1994; Tognini-Bonelli 2001).

It should be noted that discourse and corpus analysis are not mutually exclusive techniques. In many respects, a corpus is simply a composition of discourses. Correspondingly, researchers who engage in discourse analysis may decide to create their own corpora. If one wanted to draw some kind of distinction between the two approaches, one could say that while discourse analysis may, in certain cases, focus on a single speech event or a single written text (or entail an intertextual analysis of a small number of texts), corpus-based research typically deals with enormous amounts of data that may cover different contexts and genres, although there is growing interest in small corpus studies and corpora for specific purposes (see Ghadessy et al. 2001 and Waugh et al. this volume).

One question that naturally arises when contemplating the idea of working with authentic discourse concerns data access. Should one use already existing corpora, or venture into the admittedly more laborious process of collecting, transcribing, and coding data? And how does one find corpora that serve one's specific research interests? Throughout this chapter, we will point to corpora available to the linguistics community and present studies that illuminate different methods of exploiting those resources, each time in response to specific research questions.

The remaining portion of this chapter has two parts: Section 2 is devoted to cognitive-functionalist perspectives on (primarily spoken) discourse, and includes a brief discussion of multimodal and sign language discourse. Section 3 gives an introduction to corpus-based research, with a focus on the utility of corpus analysis as a methodology for extracting and examining naturally-occurring language patterns. It further contains a summary of common terminology used in corpus research as well as an annotated list of some popular corpora that may be useful for empirical investigations from a cognitive perspective. Overall, we hope to inspire the reader to consider discourse and corpus analysis as one of the keys to a fuller understanding of human cognition and social behavior.

2. Language, cognition, and social interaction: Cognitive-functionalist approaches to discourse

As we have seen in the introductory section, perspectives on what discourse constitutes are manifold. The same is true for the different approaches to discourse analysis and the types of data from which one can choose. Whether the reader is interested in working with authentic usage data because she believes that language can only be studied in context, or because she wishes to complement experimental and/or introspective methods, this section is intended to demonstrate some of the ways in which established cognitive/functionalist approaches to discourse can shed light on the interplay between language use, embodied cognition, and social interaction/situatedness. Although both qualitative and quantitative methods will be discussed, the focus in this section will be on qualitative aspects and how they may be combined with quantitative analyses. The latter will be further exemplified in Section 2 which examines corpus-based research.

The theoretical angle and the type of data one decides to work with will inevitably determine the choice of methodologies. Since it is impossible to cover all cognitive and discourse perspectives on language, we limit most of our discussion to empirical work with different types of spoken discourse (mainly narratives and talk-in-interaction). Working with natural spoken language is, as we hope to show, particularly fruitful in the context of language and cognition research.

> Ordinary talk [...] occupies a special place as the kind of language that is most natural in both form and function, the kind of language humans must be designed by evolution to produce and comprehend. [...] Because conversation is the form of language least influenced by acquired skill, it provides us with the most direct and uncontaminated access to natural mental processes. (Chafe 1998: 96–97)

As will become evident in the course of this chapter, close investigations of ordinary talk have resulted in insights into cognitive/linguistic phenomena such as flow of thought, consciousness, discourse organization, information flow, attention, coherence, interactional patterns, epistemic stances, etc. In what follows, we will discuss several interrelated approaches to discourse, all of which view grammar not as a set of abstract and disembodied principles; rather, they conceive of linguistic form and structure as being directly conditioned by fundamental practices of language use in social interaction and its cognitive, semantic, and pragmatic implications. One of the underlying beliefs is that "[g]rammars code best what speakers do most" (Du Bois 2003a: 49).

2.1 Emergent Grammar

The theory of *Emergent Grammar* views linguistic structure as coming out of language use and as being shaped by prior discourses as well as by the ongoing discourse (Hopper 1998: 156). In contrast with theories of grammar that postulate a repertoire of prediscourse and *a priori* formal rules enabling the speaker to form correct sentences in a given language and to judge the grammaticality of sentences (Chomsky 1965), emergent grammar conceives of form as the result of repeatedly and routinely used items in face-to-face interactions between interlocutors in naturally occurring communicative situations. It takes into account a speaker's past experience (in terms of the forms previously used, texts previously encountered, the situations previously experienced, etc.) and the contextual complexities and social conventions of the communicative event in which utterances are purposely produced.

The reason why we are discussing this view of grammar here is that it has both a social-interactive and a cognitive-cultural component, thus lending itself to being tested on usage data by investigating the material side of what Hopper (1998: 158) called "the partial settling or sedimentation of frequently used forms into temporary subsystems." One way of doing this is to explore a certain pattern, or two contrasting patterns, of use of a specific form or expression, focusing on its distribution and its context-dependent functions (cf. Bybee & Hopper 2001). As Hopper (1998: 161) stresses, systematicity in terms of lexical and grammatical patterns "can only be described *in situ.* Emergent regularities are aggregations; they are the sediment of frequency." Here, empirical work with larger amounts of spoken data is particularly fruitful as it may reveal how discursive strategies are employed online and under real world conditions. It comes as no surprise that dialogue, in Bakhtin's (1986) sense, is central to this view of grammar which attributes considerable weight to the fact that in a speech event, linguistic forms are constantly adapted and relativized according to cognitive constraints, the speaker's communicative goals, the hearer's reactions, as well as to the context (genre, register, familiarity with the topic, etc.). Grammar, in this view, is not a commodity that a speaker can simply use. On the contrary,

> [g]rammar, understood as meaningful repetition, is thus distributed among the various participants in a collaborative act of communication (Fox 1994; Goodwin 1979). It is also distributed among different genres of speech and among different registers (degrees of formality). Grammar is, in other words, not uniform, but relative to context.
>
> (Hopper 1998: 162)

Although individual characteristics of the speakers (such as age, gender, education, etc.) are accounted for, particular attention is given to cognitive/linguistic routines. Notably, emphasis is placed on prefabricated combinations of units with the tendency to reoccur in certain configurations in certain contexts, such as 'I'm *gonna* go' and 'I *gotta* go' instead of 'I'm *going to* go' and 'I *have to* go.' Other examples for such pre-patterned forms, or chunks, are frequently used (and thus cognitively entrenched idioms) clichés, formulaic expressions, transitions, openings, closures, greetings, requests, favored clause types, etc. (Hopper 1998: 166). One particularly interesting aspect we would like to point out is that in this view of grammar the time factor resides in more than one way. On the one hand, Emergent Grammar is built on the recognition that historically, the forms we use today have been handed down to us from previous generations of language users; on the other hand, the grammar of a language is seen as being in a state of constant flux, due to both the dynamics of language use within a speech community and also across speech communities (Hopper 1998: 172). We cannot go any further into issues of grammaticalization here; the interested reader is referred to Bybee (2003).

It is important to point out that work conducted under one of these functionalist perspectives often encompasses questions and methods from various neighboring approaches. What these perspectives seem to have in common is their desire to offer, in one way or another, a window into the interplay between conceptual, semantic/lexical, and morphological/syntactic structure as manifested in actual language use beyond the level of the sentence, either in routinized uses that, over time, have materialized in patterns, or in new trends that start to surface and gain stability in a given context. Structure, or grammar, here, is not understood as a static phenomenon, but instead as the outcome of a dynamic process – also called 'structuration' (Giddens 1984; cf. Hopper 1998: 158) – reflecting embodied patterns of experience, communicative goals, and what interlocutors esteem to be useful and appropriate according to the social and institutional settings in which they find themselves. On the basis of these insights, functionalists have claimed that both discourse location and discourse context can become predictors of grammatical choices (Hughes, Carter, & McCarthy 1995; Tao 2003b). In what follows, we will examine some of the closely related functionalist approaches in more detail, stressing each time the ways in which they offer insights into the intricate relationship between language use and cognition. As space is limited here, we trust that the reader will find productive ways to incorporate some of the research ideas and methods presented here into her or his perspective and work, including combining it with other observational and/or experimental methods.

2.2 Discourse and Grammar

Linguists working within the *Discourse and Grammar* framework are interested in "showing how fixed structures crystallize out of dynamic configurations of form and function in discourse" (Du Bois 2003a: 1). A vast array of cross-linguistic investigations into salient organizational principles in discourse undertaken so far has contributed considerably to a better understanding of the nature of speech genres and communicative behavior in general. Since this line of research makes use of both spoken and written corpora, and typically combines quantitative and qualitative methods, there are research questions and methodologies that overlap with those used in corpus-based research (cf. Barlow & Kemmer 2000; Biber et al. 1998; Partington 1998; Tognini-Bonelli 2001; cf. Section 2.2.1 of this chapter) and conversation analysts (cf. Section 1.3). The work cited here involves typologically different languages and may inspire the reader to undertake her or his own data-driven investigations into, for example, the pragmatics of transitivity (Du Bois 2003a; Hopper & Thompson 1980), ergativity (Du Bois 1987), turn-taking strategies (Goodwin 1979; Tao 1996, 2003b), repairs (Ford, Fox, & Thompson 2003; Fox, Hayashi, & Jasperson 1996), topic continuity (Givón 1987), information flow/packaging and consciousness (Chafe 1987, 1994, 1998), discourse markers (Lakoff 2001; Schiffrin 1987), personal pronouns (Fonseca-Greber & Waugh 2003; Waugh et al. this volume), or the relative frequency and distribution of (lexical/grammatical) forms in a given context or across contexts (Bybee & Hopper 2001).

One way to begin considering original research in this domain is to compare two languages with regard to the intricate relationship between emergent grammatical patterns and discourse function, thus testing the claim that "well-grounded patterns in discourse have the power to shape the very foundations of grammar" (Du Bois 2003a: 82). When comparing two or more languages, one needs to make sure to collect and work with data representing similar speech genres (narratives, ordinary conversation, job interviews, etc.), given that pragmatic factors heavily influence the grammatical choices speakers make. It is also advised to focus the analysis on a specific grammatical aspect that, in the case of comparative work, both languages exhibit in a way that makes a contrastive analysis worthwhile.

Let us now turn to some examples of such well-defined investigations, taken from research on Preferred Argument Structure carried out by Du Bois and colleagues (Du Bois 2003b; see also contributions in Du Bois et al. 2003 and Tao 2003a). Acknowledging the dominant role the verb is often afforded in the grammatical structuring of a clause, the hypothesis is that there are statistical tendencies regarding how the argument structure of a given verb is realized in particular discourse contexts. In every instance, the employment of a given verb triggers expectations regarding a matching configuration of nominal roles. For example, the English two-place verb 'name', which syntactically involves a subject (the person who names) and an object (the person or thing being named) can theoretically be realized with different combinations of lexical noun phrases and/or personal pronouns. The question that arises here centers around why these elements get realized the way they do in particular moments in a discourse, and why there are instances where, for a three-place predicate such as 'give,' only two argument roles are overtly realized (e.g., the giver and the object given are referred to but not the recipient of the object). Evidence from discourse has shown that contrary to common assumptions, three-place predicates rarely occur with all three functions realized with full lexical noun phrases (Schuetze-Coburn 1987). Other closely related issues concern discourse motivations behind transitive and intransitive uses of verbs in specific contexts (Du Bois 1987, 2003b; Hopper & Thompson 1980). The (preferred) form that arguments actually take in certain environments depends on both cognitive and pragmatic pressures:

> The grammatical realization of arguments in a clause does not take place in a functional vacuum. It is tied to cognitive and pragmatic factors like information management, which influence the realization of arguments as lexical or pronominal [...]. Corresponding to the grammatical contrast between lexical noun phrase and pronoun is the pragmatic contrast whereby the fuller form tends to be used for referents that are less cognitively accessible (Ariel 1990, 2001; Chafe 1987, 1994), while the reduced forms are used for more accessible referents. (Du Bois 2003a: 37)

We cannot go into depth here, but we will provide an idea of the sort of insights this kind of empirical work can yield. The data for the study discussed here was taken from *The Santa Barbara Corpus of Spoken English* which consists of spontaneous spoken discourse; for details about corpus design, transcription, annotation, etc., the reader is referred to

Du Bois (2000, 2003c). Du Bois defines the Preferred Argument Structure, which is regarded as a fundamental, cross-linguistically attested type of pattern at the intersection of grammar and discourse, as follows:

> Preferred Argument Structure represents neither a discourse structure nor a syntactic structure per se, but a preference in discourse for a particular syntactic configuration of linguistic elements, both grammatical and pragmatic. Roughly, the claim is that in spontaneous discourse, the distribution of nominal referential forms (such as full lexical noun phrases or pronouns) across the various syntactic positions (subject, object, oblique) is systematically skewed. Speakers freely realize full lexical noun phrases in intransitive subject position or transitive object position, but strongly avoid them in transitive subject position. In a pragmatic parallel to this, new information (typically expressed by full lexical noun phrases) freely appears in transitive subject or transitive object roles, but not in transitive subject roles. (Du Bois 2003b: 48)

Let us consider some instantiations of these principles. Although we will continue to draw on Du Bois' (2003a, 2003b) discussion of English data, we will not be able to render the work in all its complexity and scope. In accordance with Du Bois (2003a), discourse excerpts are presented in the form of intonation units following the transcription conventions established by Du Bois et al. (1993).

One-place predicates are intransitive verbs such as 'arrive' that only take one core argument, typically a subject (S) that can be realized as, more or less, substantial noun phrases or personal pronouns. The following examples are taken from Du Bois (2003a: 59):

(1) {CONCEPT 416}
MARILYN: . . **The fish** are running.
S

(2) {DOLLARS 629}
DAN: . . **policies** change

As mentioned above, two-place predicates take two core arguments. Examples are transitive verbs that invoke a transitive subject (also called the A role) and its direct object (also called the O role) such as 'enjoy,' 'eat,' 'write,' etc. The discourse excerpts given below reflect the tendency that out of the two arguments, a transitive verb takes only one which is realized as a full noun. Du Bois (1987) termed this the *One Lexical Argument Contraint*: "Avoid more than one lexical core argument" (Du Bois 2003a: 60).

(3) {HOWARDS 1102}
JANICE: (H) But I enjoyed . . . **the movie.**
A O

(4) {ZERO 489}
NATHAN: . . . She's eating **that bu=g.** (Du Bois 2003b: 59)

As we can see in (3) and (4), the subject roles are invoked by personal pronouns, whereas the object roles are realized as full noun phrases. This mirrors the empirically-founded observation that lexical nouns do not occur with the same likelihood in all roles; rather, pronouns are used much more frequently than full nouns to fill the A role (transitive subject). These observations prompted Du Bois to formulate the *Non-Lexical A Constraint*: "Avoid lexical A." One needs to keep in mind that these are soft constraints and not grammatical rules. Finally, three-place predicates, such as ditransitive verbs (give, tell, show), entail an additional argument, i.e., an indirect object position (I). Interestingly, the prevailing discourse pattern indicates that only one of the three argument positions tend to carry full noun phrases (Schuetze-Coburn 1987).

(5) {LAMBADA 1126}
MILES: I told you **that story**,
A O I

(6) {HOUSEHOLD 1245}
RON: Yeah she showed me **all that stuff.** (Du Bois 2003b: 61)

Importantly, these findings allow for predictions about where in a clause cognitively demanding tasks (such as introducing a new referent in the discourse) and non-demanding tasks (such as referring to an already introduced referent) occur. Whereas the former appears to be confined to specific environments, the latter may occur almost anywhere in a clause (Du Bois 2003a: 58). One of the major insights provided by this line of research is that the chief guiding principle behind preferred argument realization seems to be cognitive cost. Recurrent patterns observed in large bodies of authentic discourse data stemming from geographically and typologically diverse languages "point to a systematic exploitation of syntactic structure as a frame for organizing and managing cognitive cost in speech production and understanding" (Du Bois 2003a: 81). Referential-pragmatic factors lead to linguistic choices such as the ones illustrated above, which can point not only to stages of information processing on the side of the speaker, but also to the speaker's assumptions about the cognitive effort needed by the listener to access the intended referent.

There exists robust numerical evidence for these claims based on quantitative analyses of narrative discourse data in diverse languages (Du Bois 1987). While accounting for the grammatical and pragmatic differences at work in the various languages, similarities could be identified as to how Preferred Argument Structure is implemented in actual language use. In essence, the *One Lexical Argument Constraint* was found to hold in respect to data from Hebrew, Sakapultek, Papago, English, and Goonyandi (cf. Du Bois 2003b: 62, for information on the corpora used). That is, in all these languages, clauses with zero nominal argument (50–62 % of all occurrences) or only one core argument (35–53%) are much more frequent than clauses with two full noun phrases (1–7%) (Du Bois 2003b: 62). In addition, the results show clear preferences in terms of the syntactic role of lexical core arguments, or put differently, the locale where the only full noun phrase in a clause will most likely occur (see the *Non-Lexical A Constraint* discussed above). Evidence from Hebrew, Skapulek, Papago, English, Spanish, French, Brasilian Portuguese, and Japanese suggest that lexical arguments freely occur in the S role (subject of intransitive verbs; 35–58%) or in the O role (object of transitive verbs; 37–58%). However, they have a tendency to avoid the A role (subject of transitive verbs; 5–10%; Du Bois 2003b: 63). Crystallized grammatical structure then can be seen, Du Bois (2003b: 82) concluded, as "an architecture for cognitive processing, in which certain locales are predictably specialized for high- or low-cost work" (ibid.), or in other words, as "a cognitive resource available to all members of the speech community" (p. 82). Even though we could only present a glimpse of this research, we hope to have shown how pervasive such clause structure preferences are in real discourse; for a full appreciation, the reader is referred to Du Bois (1987, 2003a, b) and Du Bois et al. (2003).

2.3 Social Interaction and Grammar

In a similar vein, linguists working within the *Social Interaction and Grammar* paradigm (Ochs, Schegloff, & Thompson 1996) view grammar "as sets of practices adapted to social interaction," recognizing the need for a perspective on grammar in use that "must be both cognitively realistic as well as interactionally sensible" (Ford, Fox, & Thompson 2003: 119). In other words, people working within this framework note that conceptual and social structure must be taken into joint consideration when exploring the linguistic structure underlying contextualized communicative behavior. Here again, empirical work with discourse data can help us understand cognitive and linguistic trends. Whereas the data discussed in the previous section consisted primarily of narratives, the focus will now shift to conversational data, and thus to the interactional give-and-take with which speakers co-construct their discourses. With recourse to Edelman's work (1992), Ford, Fox, & Thompson (2003: 120/1) emphasize the fact that "the human brain is exquisitely adapted to be very good at remembering, storing, categorizing, and using routines that have proven useful for solving everyday problems; with frequent repetition, as synapses become strengthened, these routines become crystallized as habits." The idea is that engaging in certain types of talk-in-interaction (such as friendly conversations at the dinner table, phone conversations, interviews, service encounters, etc.) has "a number of grammatical consequences" (ibid., 121). According to Schegloff, who postulates an orientation towards *action* and interaction in discourse, instantiations of discourse elements, such as turn initiators and closures, in ordinary acts of talk can be regarded as "interactional achievements" (Schegloff 2001: 229ff.; see also Goodwin 1981), reflecting the consequences of the presence or absence of action for the constitution of discourse.

Not only grammatical constructions, but also, among other things, length and placement of pauses, hesitation phenomena, and listeners' feedback (in the form of minimal reactive tokens – also called back-channeling – such as *oh, yeah, mhm, uh huh*), and laughter are seen as feeding into the dynamic co-construction of a conversation (Erickson 1985; Schegloff 2001; Tao 2003b). Listening behavior, too, gives insights into cognitive processes and cultural schemata of interpretation and expectation (Erickson 1985: 305). As data artificially constructed by linguists typically do not feature otherwise naturally occurring discourse particles, strong arguments have been made in favor of empirical discourse linguistics that can demonstrate the degree to which discourse elements of all sorts of complexity have both cognitive and social significance (Chafe 1998; Ford, Fox, & Thompson 2003: 123; Schiffrin 1987).

Across these approaches to discourse and grammar, one can see efforts undertaken by the researcher to balance quantitative and qualitative methods of analysis. This is evidenced by work that combines large quantities of data with a rather detailed analysis of single episodes of discourse, as done in the conversation analysis (CA) tradition (Duranti

& Goodwin 1992; Schegloff 1987, 2001, inter alia). On the basis of transcripts of talk-in-interaction data, qualitative analyses allow for the identification of discourse strategies as they underlie longer stretches of talk (Goffman 1981; Gumperz & Berenz 1993; Schiffrin 1994; Tannen 1990). We will now look at two studies representing these approaches to language in use.

In his extensive corpus study of forms and functions of turn initiators in spoken English, Tao (2003b: 204) argues in favor of combining computer-assisted analysis of large amounts of data with micro-analytical methods in order to discern, in the case of the study cited here, the often subtle but decisive contextual factors that affect the mechanisms of turn-taking. He defines a turn initiator (p. 189) as any form that starts a new turn. Tao used two corpora of spoken American English: the Switchboard corpus (a collection of telephone conversations, cf. Godfrey, Holliman, & McDaniel 1992) and the Cambridge University Press/Cornell University Corpus (consisting of informal conversations among family members and friends). Altogether, 3000 turns were randomly selected, and a computer program searched the corpora for the first form coming after the speaker label. Here is an example from the Switchboard corpus (Tao 2003b: 196; turn initiators are underlined):

> A.93: <u>You</u> can only laugh [laughter].
> A.94: <u>Yeah</u>, you just sort of, you know, well I guess I can just humor them, you know [laughter]. At this point.
> A.95: <u>Right</u>. [laughter] well, they, I guess our age is showing when, we, we think that.
> B. 96: <u>Yeah</u>, but well, you know, I, I, I've liked a lot of the new music.

In addition to supporting Schegloff's (1996) observation that lexical forms dominate at turn beginnings, Tao's study also provides insights into the grammatical, syntactic, and pragmatic characteristics of turn initiators. He found, for example, that the majority of turn initiators are syntactically independent forms ('well', 'yeah', 'oh,' 'and,' 'right,' 'so,' 'you,' etc.) and that the determiner 'the' – which is usually taken to introduce a noun phrase at the beginning of sentences – rarely opens a turn in conversation (3% of all the 'the' forms occur in turn-initial position). According to Tao (2003b: 194), this seems to suggest that "(1) turns and sentences are two very different levels of linguistic organizations – if sentences can be shown to exist at all in spoken data; and (2) that there is a striking discrepancy between constructed samples and how turns actually occur." Moreover, by combining qualitative methods with the initially achieved quantitative results, Tao established the following four functional categories of common turn initiators: tying ('oh,' 'well,' 'but,' 'and;' connecting to other previous turns (Sacks, Schegloff, & Jefferson 1974; Schegloff 1996), assessing ('yeah,' 'so,' 'right'), explaining ('so'), and acknowledging ('mhm,' 'uh-huh,' 'okay'). The analysis of the relative frequency and co-occurrence of forms, as well as their specific location and function in the discourse structure, thus seems to be a fruitful endeavor. Referring to Schegloff (1993), Tao (2003b: 204) reminds us that "the key to the usefulness of quantitative analysis of social interaction lies in a well-defined set of comparable variables." In his study, the variables were limited to an identifiable turn position.

Using a similar approach, Tao (2003a) has offered compelling results indicating that the verb 'remember' – traditionally categorized as a transitive verb of cognition that can take a range of complement objects (as in 'John remembered to turn the faucet off', 'I remember turning off the faucet', and 'John remembered that he had left the faucet on') – exhibits highly recurrent discourse behavior that challenges received views based on constructed examples and imagined syntactic behavior and meaning. With the findings that complements are rare in the data and that 'remember' is often used with zero subjects and/or zero objects in discourse, Tao concluded that the verb 'remember' can rather be seen as an interactional marker of epistemic stance (as in 'remember... you're gonna spend the rest of your life with me'). It seems safe to say that such balanced approaches to face-to-face interaction have the potential to challenge other established ideas about language, grammar, and cognition that are mostly rooted in idealized data (see also, among others, Tognini-Bonelli (2001: 15ff.) for corpus evidence for expected and surprising uses of 'any'). The work represented in this chapter is a step toward a more ecologically sound understanding of the contextual and functional motivations of linguistic form; however, as a lot more work is needed in this domain, we hope that the reader will feel compelled to engage in similar empirical investigations.

2.4 Different degrees of contextualization: Discourse pragmatics and ethnography

Here, a link can be made to other discourse pragmatic approaches which not only share beliefs about the nature of language, as the ones discussed above, but also about adequate methods of inquiry. The different empirical traditions that can be more or less subsumed under the umbrella of *discourse pragmatics*, vary, however, in the degree to which they include cultural and social factors in the analysis of usage data.

As demonstrated in the chapter by Waugh et al. (this volume) on multiple paths to an integrated analysis of authentic spoken discourse, one must often integrate several empirical methods in order to arrive at a full understanding of why speakers employ the words and structures they do in a given language, in a given act of communication (i.e., speech event), and in a given socio-cultural context (i.e., speech community; for an explanation of these basic terms see Waugh et al. this volume). The authors exemplify the different stages of data-driven investigations into linguistic phenomena, such as the use of personal pronouns in different varieties of modern spoken French, as well as accommodation strategies and other interactional practices among native and non-native speakers of English in university settings. It is important to realize that "data-driven" does not equal "data-based," since in data-driven work the data are not just the material basis for the analysis, or just used to test or exemplify theories formulated prior to the advent of large corpora. Rather, the data actually guide the course of analytical processes and theory building (for a detailed discussion of corpus-based vs. corpus-driven approaches, see Tognini-Bonelli 2001: 65ff.). Especially in the course of the first study presented in the chapter (friendly conversations in French), it becomes evident how some of the first findings regarding usage patterns may prompt the researcher to adjust the initially adopted methodology, depending, of course, on the nature of the original research questions and those questions that may arise from the corpus material. Working from within a single empirical tradition can entail the risk of limiting the scope of both the analysis and the results. It is thus recommended to be open to consulting a variety of compatible frameworks, compatible in the sense that they share fundamental beliefs about the nature of language and grammar and that their methodologies are apt to be applied to the same type of data. Therefore, it is the researcher's decision whether or not to focus primarily on linguistic issues such as lexical choices, discourse markers, or grammatical structures (thus staying within the confines of the discourse at hand while trying not to let her or his own perspective play a role in the interpretation). Waugh et al. make the point that the linguistic strategies speakers employ cannot be sufficiently explained without considering their respective age, gender, cultural origin, level of education, socio-economical status, and the stance the participants assume in a given speech event (which is always anchored in a particular society and/or institutional context). Given their research interests and data, the authors deemed it appropriate to combine methods stemming from conversation analysis (CA) and ethnography of communication into *culturally contextualized conversation analysis* (CCCA; cf. Moerman 1988), combined with other approaches (for work on written genres done from a discourse pragmatics perspective, see Fleischman & Waugh (1991) and Waugh (1995).

Ethnography of communication, especially as practiced in sociolinguistics and anthropological linguistics (Duranti 1997; Duranti & Goodwin 1992; Gumperz 2001; Hanks 1996; Hymes 1961, inter alia), entails the detailed description of the cultural, social, and institutional conditions under which individuals function in their communities, and is a particularly useful methodology for linguists conducting fieldwork. It is believed that the rules governing language use can be established by a combination of systematic participant observation, analysis of spontaneous language, and interviews with the speakers. The way in which discourse proceeds is regarded as demonstrating how social identities are negotiated (Gumperz 1982; Goodwin 1981; Labov 1972).

Of particular interest to the cognitive linguist is the recently established sub-discipline *cognitive ethnography*. Integrating distributed cognition with cognitive semantics (notably, *Conceptual integration theory* (CIT), Fauconnier & Turner 2002), Williams (2004) investigates the interplay between material structures and conceptual operations in situated cognitive activity, namely how children learn how to tell the time with a clock. The complexity of this cognitive task is reflected by the all-encompassing methodology: the gathered data comprise participant observation, audio-video recordings (elementary students and teachers during math lessons), interviews, and artifact analysis. Examining image schemas, conceptual mappings, and material anchors (Hutchins 2002), Williams demonstrates how speech, gesture, and artifact structure prompt, guide, and ground blends in interaction, thereby driving the construction of meaning in the unfolding discourse. Work like this suggests that one of the most promising areas of conceptual integration theory, also known as mental space theory or conceptual blending theory, is its application to discourse data (cf. Hougaard 2005; Pascual 2002; Parrill & Sweetser 2004 discuss its merits with regard to gesture analysis and transcription; see also Sweetser this volume). In essence, this theory offers a model of human information integration, i.e., of creative online meaning construction, involving "a set of operations for combining dynamic cognitive models in a network of 'mental spaces' (Fauconnier 1994), or partitions of speakers' mental representations" (Coulson & Oakley 2000: 176; see also Grady, Oakley, & Coulson 1999).

2.5 Language and the body: Multimodal communication and sign language discourse

A considerable number of conversation analysts have come to incorporate speech-accompanying bodily actions, such as gaze and gesture, in their analysis of communication, thus bringing to light how the human body, cognition, and language interact in socially and ecologically grounded meaning-making processes. Research topics include the function of communicative body movements in lexical searches, information processing, discourse structuring, turn constructions, speaker selection/floor management, aspect, deixis, and more global topics such as multimodal teaching and learning (e.g., Clark 1996; Duncan 2003; Duranti 1997; Furuyama 2000; Goldin-Meadow 2003; Goodwin 1981, 2001; Kendon 1990, 1995, 2000, 2004; Kita 2003; Kress et al. 2001; McNeill, 1992, 2000, 2003; Müller 1998, 2004; Smith 2003; Streeck 1993, 1994; Tabensky 2001). This research illuminates the degree to which human behavior seems to be patterned on a number of planes (conceptual structure, grammar, prosody, bodily movements, etc.) and how social actors creatively use their cognitive/semiotic resources in the construction and expression of their identity, both as an individual and group member.

Over the last decade, cognitive linguists have significantly contributed to the growing body of interdisciplinary research on gesture. Investigations into the logic and use of spontaneous co-speech gesture have not only enhanced our understanding of situated, distributed cognition, but have also resulted in additional evidence for conceptual metaphor. Gestural representations of abstract concepts and structures may reveal source domain information not necessarily captured by concurrent verbal expression. The manual modality has been found to be particularly apt at depicting spatial, dynamic, and sensory-motor properties of mental imagery and conceptual processes, thus supporting theories of embodied semantics (e.g., Bouvet 2001; Calbris 2003; Cienki 1998, 2005; Cienki & Müller fc.; Gibbs 2003; McNeill 1992, 2005; Mittelberg 2006, fc.; Müller 1998, 2004; Nuñez 2004; Nuñez & Sweetser 2006; Parrill & Sweetser 2004; Sweetser 1998; Taub 2001). Among other things, gestures may provide a window to the on-line processes of (figurative) thought, discourse management, and the use of space as a tool that makes cognition and knowledge interactively available. For anyone who is interested in exploring the multidimensionality of meaning construal in authentic discourse, working with multimodal data may be a promising enterprise to consider. We will stop here, as two of the more specific chapters in this volume are devoted to gesture research and methodological issues arising from working with speech and gesture data. Sweetser (this volume) applies mental space theory to co-speech gesture, and Mittelberg (this volume) walks the reader through one of many possible ways of eliciting, transcribing, coding, and analyzing authentic speech-gesture data.

Given that gesture and signed languages use the same media of expression, namely hands and space, there has been a lot of cross-fertilization between the gesture research mentioned above and cognitive approaches to American Sign Language (ASL) (Dudis

2004; Liddell 1998, 2003; Taub 2001; P. Wilcox 2000, 2004; S. Wilcox 2004) and French sign language (Bouvet 1997). Of course, signed languages are, compared to spontaneous gestures, highly conventionalized semiotic systems with elaborate mechanisms of sign constitution, morphology, syntax, discourse structure, etc. Yet, together these lines of research have offered support for the idea that visuo-dynamic modalities can indeed provide evidence for central claims made in cognitive linguistics, and especially in cognitive metaphor theory (Johnson 1987; Lakoff & Johnson 1980, 1999; Lakoff 1987, 1993; Sweetser 1990).

Since Wilcox and Morford discuss in their chapter (this volume) the ways in which cognitive linguistics has brought new energy to the study of signed languages, we will not go into further detail here. The authors present a variety of theoretical perspectives and empirical methods that have been adopted by sign language researchers working within cognitive frameworks. Importantly, these tools allow for the investigation of phenomena that are at the center of interest within the cognitive linguistics paradigm: grammaticization, polysemy, iconicity, metaphor, metonymy, and so forth. In their discussion of different ways of dealing with the challenges of documenting, transcribing, and analyzing sign language discourse, the authors stress the merits of integrated approaches that combine several methodologies. We can thus notice a trend toward developing multimethodological approaches to discourse, a procedure that is also strongly recommended by Waugh et al. (this volume) and Grondelaers et al. (this volume).

2.6 Mediated situatedness: A look at recent accounts of metaphor in (public) written discourse

While this chapter has so far focused on spoken discourse, we would like to point the reader to recent empirical work on written discourse carried out from a cognitive linguistics perspective. We will only be able to touch on a couple of aspects; the focus will be on methodological aspects analyzing conceptual metaphor in public discourse.

As is the case with spoken language, the question of genre is also central with regard to pragmatic factors that shape written discourse. From the beginning of the cognitive linguistics enterprise, literary texts have served as rich sources for linguistic metaphorical expressions (Lakoff & Turner 1989; Turner 1987). This kind of systematic metaphor analysis, essentially consisting of identifying metaphorical concepts and inferences, has later been complemented by investigations into psychological aspects of processing and understanding metaphor in literature and journalism (cf. Gibbs 1994; Steen 1994). Such empirical work exemplifies, as did the work on spoken discourse discussed above, "the interaction between theory and data" (Steen 1994: 242; see also Steen 1999 for a concise description of a step-by-step metaphor analysis; Cameron & Low 1999; Glasbey, Barnden, Lee, & Wallington 2002; Kövecses 2002; Lakoff 1987; Steen 2004b).

Here, we would like to draw attention to recent work on sociocultural aspects of metaphor use, or metaphor practices, in journalistic discourse. In his study of national identity construction in Russian and German public discourse, Zinken (2004) argues that quantitative analyses can enhance comparability in large-scale cross-cultural and cross-linguistic research, and provide the foundation for local, qualitatively-oriented analyses of metaphor (or any other discourse feature). Focusing on conceptualizations of major political transformations (the German 'Wende' discourse – 'Wende' signifies 'turn' and stands for reunification – and the Russian 'perestroika,' i.e., 'rebuilding,' discourse), Zinken carried out a three-step, corpus-driven analysis of newspaper texts: 1) through a frequency analysis, metaphorical models are grouped into clusters; 2) the cognitive stability of such clusters is assessed; and 3) metaphor models and constellations of models are assessed as parts of the group's discursive practice (Zinken 2004: 2–3). Results indicate that the most frequently exploited source domains are, in the case of the German texts, SPACE, PERSONIFICATION, MOTION, OBJECT, WAR, ARCHITECTURE, which led the author to identify two fundamental perspectives that seem to frame these society-transforming events. The first is the perspective of an active agent exerting influence on real-world events (reflected by metaphor models such as ARCHITECTURE, MOTION, WAR, etc.); the second is the perspective of the passive subject lacking control over things (reflected by metaphor models such as ORGANISM, PLANT, WEATHER, etc.). In both cultures, the former is more common than the latter; however, the German discourse showed an especially strong tendency to frame socio-political change from the position of the passive observer (Zinken 2004: 11). Such observations offer valuable insights into the ways in which cultural-historically rooted perceptions and imaginations may crystallize in the course of such methodologically complex metaphor analyses.

Taking these issues into the larger context of discourse analysis and cognitive metaphor theory, Zinken, Hellsten, and Nerlich (fc.) propose the notion of 'discourse metaphor' which they define as "a relatively stable metaphorical projection that functions as a key framing device within a particular discourse over a certain period of time." The authors make a distinction between primary metaphors (Grady & Johnson 2002) and discourse metaphors and postulate that the latter indeed have a high degree of phenomenological salience: "they use knowledge associated with basic level concepts; they evolve in social interaction; and they are firmly linked to cultural scripts and stereotypes" (p. 13).

In accordance with Sinha's (1999) notion of 'dual grounding' of human cognition in both biology and culture, Zinken, Hellsten, and Nerlich argue in favor of an approach to metaphor that accounts for the intricate relation between embodiment of cognition and its sociocultural situatedness. Consideration of the cultural origins of cognition that have shaped and still shape the processes of conceptualization and 'enculturation' seems to be a priority on their research agenda. Metaphors, the authors further maintain, transport the images, feelings, values, thought patterns, etc. entrenched in our cultures (cf. Dirven, Polzenhagen, & Wolf fc.; Dirven, Wolf, & Polzenhagen fc.; Kövecses 2005; Nerlich, Hamilton, & Rowe 2002; for introduction to the notions 'frame,' 'script', etc. see Croft & Cruse 2004; for more discourse and/or corpus studies on related issues see Deignan 1999; Musolff 2004; Settekorn 2001a/b).

In light of the different efforts to understand cognitive and semantic structure through the study of language in context, we can say that discourse (no matter what form it takes) mediates biological, social, ideological, political, and cultural structures and schemata. As work within the field of social semiotics and critical discourse analysis (CDA) has shown, semiotic practices, i.e., the use of specific semiotic resources (speech, gesture, writing, images, etc.) in social situations, are deeply seeded in ideologically and historically conditioned representational conventions (Fairclough 1995a/b; Hodge & Kress 1988;

Kress & van Leeuwen 1996; Settekorn 2003; van Dijk 1997a/b). According to van Dijk (2001: 352), "[c]ritical discourse analysis (CDA) is a type of discourse analytical research that primarily studies the way social power, abuse, dominance, and inequality are enacted, reproduced, and resisted by text and talk in the social and political context." Fairclough and Wodak (1997: 271–80) summarize the multi-faceted dimensions of discourse and critical discourse analysis as follows (van Dijk 2001: 353):

- CDA addresses social problems;
- Power relations are discursive;
- Discourse constitutes society and culture;
- Discourse does ideological work;
- Discourse is historical;
- The link between text and society is mediated;
- Discourse analysis is interpretative and explanatory;
- Discourse is a form of social action.

Given the spectrum of linguistic and contextual phenomena we have sketched in this section, it becomes evident that the researcher can decide to what degree she wishes to narrow down the scope of her analysis and relate the cognitive aspects of language she is interested in with issues of pragmatic function and sociocultural situatedness.

3. Introduction to corpus-based approaches

The analysis of naturally-occurring language data, as represented in the prodigious number of corpora that currently exist, can provide most language scientists, no matter what sub-discipline one works in, with a vast amount of valuable information. In light of this importance, the purpose of this present section is twofold. First, we aim to provide a brief (yet still informative) introduction to the utility of corpus analysis as a methodology for extracting and examining naturally occurring language patterns. This introduction will be accompanied by a summary of common terminology one might encounter when working with, or reading about, corpora, and should help to prepare the reader for the discussion of more thorough and complex corpus-based work in the chapter by Grondelaers, Geeraerts, and Speelman (this volume) and about combining corpus-based work with other empirical approaches in the chapter by Waugh et al (this volume). Examples of corpus-based work conducted within the cognitive linguistics framework should serve to illustrate the diverse set of situations under which corpus-based data can facilitate an empirical investigation, and subsequent illumination, of cognitive linguistic theory. Secondly, we aim to provide an annotated list of some popular corpora that may be useful for empirical investigations into naturally occurring language patterns.

3.1 Corpus-based research for the cognitive linguist

It should be noted, first and foremost, that this section is intended only as an introduction to the large area of corpus-based research. No single book chapter could possibly cover all of the topics related to the examination of corpora. Indeed, many complete volumes have been written on the topic of corpus analysis (see Biber, Conrad, & Reppen 1998; McEnery & Wilson 1996; and Meyer 2002, for assiduous and succinct introductions to corpus-based research. Both Partington 1998, and Hunston 2002, present excellent examples of corpus-based approaches to applied linguistics. Also, see Stubbs 1996, for a more thorough treatment of the corpus-based approach, and see Botley, McEnery, & Wilson 2000, for information on multi-lingual corpora).

In attempting to enumerate a subset of the ways that corpora can be used to inform theory, the distinction between *qualitative* and *quantitative* analysis becomes useful. As noted by McEnery and Wilson (1996), the notion of qualitative data analysis, applied to corpus-based research, encompasses attempts to determine the types of features that occur within the sample. For example, Partington (1998), working with the UP IS BETTER, DOWN IS WORSE metaphor (a variant of the GOOD IS UP, BAD IS DOWN metaphor, Lakoff & Johnson 1980), extracted all of the synonyms for UP and DOWN from a large corpus of business-related text. He subsequently examined all of the surrounding linguistic contexts in which UP, DOWN, and their synonyms occurred. Focusing on UP and its synonyms, Partington found many cases in which UP denoted something good, although he did observe some counter-examples (such as costs, inflation, and unemployment going UP). (For other examples of qualitative analysis of corpora, see Waugh et al., this volume.) The qualitative approach can be contrasted with a quantitative approach, whereby information regarding the *prevalence* of some feature of interest is extracted form the corpus. In other words, quantitative corpus-based approaches involve, for the most part, extracting the frequency of occurrence of a feature to be studied. In the Partington example above, if someone were to extract each sentence using UP (or one of its synonyms), and then count the number of times those words were used to denote something BETTER versus something WORSE, they would have a measure of the relative frequency with which UP is used positively versus negatively.

Although qualitative corpus analyses are beneficial in that they can facilitate a descriptive understanding of the nature of a target language as represented by a given language sample, they are often tedious, especially when done on large corpora, thus requiring a substantial time commitment from the researcher. Hence, qualitative corpus analyses tend to be done on small corpora (see Waugh et al., this volume); or, they are often a starting point for the researcher since they serve to illuminate the types of language use in a population of interest, and as a result, help the researcher to develop hypotheses that can be tested through further corpus analyses and the implementation of experimental methodologies and statistical measures. Given that Waugh et al. provide examples of qualitative corpus analysis, the focus of this current section will be on the quantitative analyses of corpora. Indeed, quantitative analyses (usually in the form of frequency counts) are much more prevalent within published corpus-based research, and serve the purpose of providing a wealth of usage-based data on naturally occurring language patterns.

Gries (2003), for example, used a corpus to help in the identification of prototypical instances of syntactic constructions. As noted by Gries, although human judgments of semantic category prototypes (like objects and colors) may be useful, human judgments of prototypes of syntactic constructions are likely to be highly inconsistent and of little use in identifying prototypical syntactic category exemplars. In order to investigate construction prototypicality, Gries extracted a series of verbs from the British National Corpus that allow for the dative alternation. Basically, some verbs (e.g., *gave*) allow for either two NPs to follow them, as in the ditransitive construction *Sally gave him the check*, or allow for a preposition to signal the indirect object, as in *Sally gave the check to him*. Gries then used frequency information from the corpus to determine the probability with which the verb would take the ditransitive versus the prepositional indirect object. The probability information, in conjunction with information from other sources, was submitted to multivariate statistical analyses (the details are not important here) which then yielded a measure of how typical the ditransitive or prepositional construction was of other constructions within the same category. Through the quantitative analysis of naturally occurring language data present within a corpus, Gries was able to extract the information necessary to design a new measure of construction category typicality.

Through the Partington (1998) and Gries (2003) examples above, one can begin to see the wealth of ways that information extracted from corpora can help the researcher to identify and explain aspects of naturally occurring language. Interestingly, corpora have also been used as a tool for extracting frequency information to help explain patterns of reading times in human language processing experiments. Indeed, frequency effects have been commonly observed within the sentence processing literature. More specifically, everything else being equal, words or structures that are very frequent tend to be processed more quickly than ones that are very infrequent in naturally occurring language.

(7) a. I know that the desert **trains** *could resupply the camp.*
b. I know that the desert **trains** *soldiers to be tough.*

A classic example of a situation where frequency information might matter can be seen in the sentence fragment *I know that the desert trains...* (Frazier & Rayner 1987; MacDonald 1993), in which the lexical ambiguity of the homonym *trains* (which is ambiguous as to its grammatical category and meaning) introduces a syntactic ambiguity with respect to the continuation of the sentence whereby a noun reading (of *trains*) would lead to the expectation of an upcoming verb (as in, ... *could resupply the camps.* (7a)), and a verb reading would result in the expectation of some type of complement (as in ...*soldiers to be tough.* (7b)).

MacDonald (1993) speculated that when an ambiguity arose as a result of a noun/verb homonym (like *trains*), the frequency that the homonym was used as a noun versus as a verb might influence, in part, whether or not people interpreted the homonym as a noun or as a verb. To investigate this effect, approximately 300 sentences containing each of the N/V homonyms were extracted from the Wall Street Journal corpus, and were coded for how often the homonym was used as a noun versus as a verb. As predicted, MacDonald found that there was a relationship between frequency of usage as a noun versus as a verb and the amount of time it took participants to read the disambiguating words (the completion options, or words occurring after the homonym, italicized in example (7)). When the homonym was used more frequently as a noun, and the ambiguity was resolved in accordance with the noun interpretation (as in *could resupply the camps.* (7a)), participants tended to be faster in reading the noun resolutions than they were the verb resolutions, and vice versa for the verb-resolved sentences.

Of course, in the language processing example above, many other factors could have influenced the reading time patterns described. However, the example does help to demonstrate two potential uses of corpora for people interested in conducting experiments. First, in light of frequency effects alluded to above, extracting frequency information from corpora can help researchers to better understand the nature, and perhaps even the cause, of reaction time patterns in experiments involving language processing. Frequency is an important variable that can influence how people process lexical and syntactic information. Additionally, however, frequency information can also help the researcher to exert better experimental control over materials they present to participants when conducting experiments involving language. If we know that frequency tends to influence reading times, and we are interested in the way that some other variable could influence reading times on the same set of materials, it is probably best to control for frequency (by using information extracted from a corpus to make sure that one condition is not more or less frequent than the other) in order to try to isolate just the other variable of interest. In other words, by controlling for frequency information, one can be more certain that any observed effect of a different variable is due to that variable, and not to a confounding frequency effect.

Above, we have presented a small subset of the many possible ways to use a corpus. Whether one aims to describe naturally occurring language patterns or to use statistical/frequency information extracted from a corpus to better design an experiment or help in accounting for results of an experiment, corpus analysis is a helpful skill for any researcher in the language sciences. Given that this introduction to the utility of the corpus has been brief, the reader is encouraged to examine some of the many existing volumes on corpus-based analysis (some cited at the beginning of this section) Also, Grondelaers et al. (this volume) should be consulted for an instructive example of how one may use a corpus, quantitatively, to test operationalized hypotheses, but then also, to develop new hypotheses that can then be tested with experimental methodology.

3.2 Terminology

The purpose of this section is to introduce the reader to some terms that commonly appear in conjunction with corpus-based research. These terms are likely to be encountered when reading about corpus-based research, and indeed, many of them are used in the latter portions of this chapter, and in later chapters within this present volume.

One important distinction inherent within the corpus-based approach is the distinction between tagged and untagged corpora. The concept of tagging can be thought of as somewhat analogous to annotation. An untagged corpus is one in which the corpus entries are not supplemented with any additional information. That is, an untagged corpus is a corpus with text appearing in its raw form. These corpora provide the researcher with little information regarding linguistic elements such as part-of-speech, pronunciation, etc. As such, untagged corpora are of a substantially reduced value to the language researcher interested in linguistically oriented phenomena. Tagged corpora, on the other hand, include, within a word's corpus entry, additional linguistic information. Part-of-speech tags, for example, are quite common. As such, when a word is identified in a corpus, one can discern its part-of-speech based on an extra piece of information stored along with the word. One can also utilize tags in order to extract words of a certain category. If one were interested, for example, in the word *train*(*s*) used as a noun, one could search the corpus for the tag '*noun*' along with *train*(*s*).

As noted above, corpora are often used in order to extract various types of frequency information from samples of naturally occurring language. The nature of these frequency counts is quite varied, ranging from simple to complex. At the broadest level, there exists an important distinction between type and token frequency. Type frequency refers to a count of the number of types of something that occur within the corpus (it can been seen as a measure of what exists within a corpus). As a basic example, consider a situation where a researcher is interested in understanding the different types of grammatical categories (e.g., noun, verb, determiner, adjective, adverb, pronoun, etc.) used by three-year-old children. Using a tagged corpus, the researcher could easily find out how many different grammatical categories are used by three-year-old children. So, if the children in the corpus used only nouns, verbs, pronouns, and adjectives, the type frequency would be four. Another example would be to list the different words that are used in a corpus of economics texts as against a corpus of academic English: the former would have words (types) like *price, cost, demand, curve, firm, supply, economy*, whereas the latter would have *introduction, relevance, investigation, theory, concept*. These word types are also called lemmas, or lexemes, since a lexeme like *firm* actually stands for a number of word-forms (*firm, firms, firm's, firms'*), all of which are instantiations of the lexeme *firm*. Type frequency is often contrasted with token frequency, whereby all the instances of each type are counted. In the three-year-old child example above, each occurrence of a noun and of a verb would be counted, yielding an indication of noun use relative to verb use, within the sample. Another example would be to count the number of times (tokens of) the word *food* occurs in a given corpus of children's speech. A list of the ten most frequent words (tokens) in a general corpus of English has the following, in order: *the, of, to, and, a, in, that, is, it, for* (adapted from Hunston 2002: 4).

Many times, one might be interested not only in single words but also in knowing the number of times words occur together. There are various programs for working with corpora, called concordancing programs (or concordancers) that search a corpus "for a selected word or phrase and presents every instance of that word or phrase in the center of the computer screen, with the words that come before or after it to the left and right" (Hunston 2002: 39); a concordance is, thus, a list of all the contexts in which a given word occurs in the corpus (and is sometimes abbreviated as KWIC=key word in context). For example, a concordance for the word *interested* shows that it may be preceded by a number of different words, but is usually followed by the preposition *in* (e.g., *people who are interested in water sports, X claims to be interested in lobbying, I'm more interested in musical instruments, we've been interested in looking at alternative methods* – Hunston 2002: 9). Such concordances can be used in a variety of ways. One popular use is to look at two or three words that often occur together: *interested in*, is an example of a two-word pattern, sometimes referred as bigrams, or a collocation, or *in* is said to be a collocate of *interested* (these are only some of the terms used in this area), and the number of times such bigrams or collocations occur is referred to as the bigram or collocation frequency. There are formulas that can be used to generate a probability estimate, and thus, a bigram probability refers to the probability that two words would appear next to each other. This probability estimate takes into account, in part, the number of words contained within the corpus, and as such, yields a probability estimate given the nature of the sample in which the bigram appeared. These formulas are too detailed for the present section, and the reader is referred to Manning and Schutze (1999) for further information regarding corpus-based probability estimates.

But concordancers can also find sets of words that are longer than two or three words (e.g., *at this point in time*), idioms (e.g., *call it a day*), fixed phrases (e.g., *that's easier said than done*), etc. Other uses of concordancing and collocations are for determining the different meanings that a word (or set of words) has in different contexts and to find out which of the meanings is typical of that word (most frequent); for seeing how the different meanings of a word are related to its different contexts; for observing the details of how words are used. For example, the nouns *advice* and *answer* are both preceded by verbs like *get, give, want, offer* and are followed by *as to*, but *answer* is typically preceded by adjectives like *definitive, clear, exact, persuasive*, and in a context that indicates either "that a clear answer is not available, or that to give a clear answer is difficult or unexpected" (Hunston 2002: 51). In other words, the context in which *answer* is used is often a negative one, something that might not be accessible to intuitions, and thus this brings up one of the important uses of corpora: to observe hidden meanings, to find patterns of contextualized meanings that are not obvious to the user or to the researcher.

3.3 Annotated list of popular corpora, plus a brief caveat

An impressive number of corpora currently exist, spanning many different languages, modalities (i.e. spoken vs. written language), and targeted genres. Indeed, a quick websearch reveals corpora with titles illustrating the varied nature of existing corpora. From *The Corpus of Spoken Israeli Hebrew* to *The Corpus of Estonian Written Texts* to *The International Corpus of English*, it should be evident that corpora of an extensively varied nature are in existence. Below, we aim to provide an extremely abbreviated list of corpora that may be of particular interest to readers of this current volume. All corpora detailed below are in English, and are the corpora most widely used in cognitive linguistic and psycholinguistic research involving the details of language and language processing.

First, however, one pressing issue should be briefly addressed. With the existence of so many corpora, one may quickly ask, "Well, how do I know which one to use?" There are some important points to consider when selecting a corpus to analyze. First, in determining which corpus to use, there are a series of pragmatic issues to consider. Specifically, one would want a corpus that can be analyzed easily by a computer software package, and one that is appropriately tagged (if one were interested in syntactic information, a corpus tagged with syntactic information would probably be the appropriate choice). One also needs to consider the population they are interested in studying. Most obviously, it would not be gainful to analyze *The Wall Street Journal* corpus if one were interested in the nature of the linguistic input available to an infant.

Of additional interest, as noted by McEnery and Wilson (1996), is the issue of *representativeness*. To the degree that a corpus can be considered a sample of language from a targeted population, one would want to analyze a corpus that is *maximally representative* of the population of interest. Assessing the concept of maximal representation is not easy given that many factors can influence the representativeness of the sample. In order to assess representativeness, at the most surface level, one might consider how the corpus was constructed. Specifically, if one were interested in a representative sample of adult American English, most broadly defined, one would hope to identify a corpus that includes samples of both written and spoken English, taken from many different types of sources (newspapers, magazines, transcriptions of language spoken on television and radio shows, etc.), and spanning many different genres (see McEnery & Wilson 1996 for a more thorough treatment of representativeness).

Here, we present a list of five popular corpora with the goal of providing the reader with a list of potential starting points.

3.3.1 *Child Language Data Exchange System* (*CHILDES*) (*MacWhinney 2000*)

CHILDES is an invaluable resource for anyone interested in either first or second language acquisition. CHILDES is an extensive database dedicated to documenting many aspects of child language use such as caregiver-child interactions, child-directed speech, and so on. The CHILDES database is a collection of corpora comprising an expansive collection of language samples. Although many of the corpora are in English, data from approximately 20 other languages are also represented. The corpora contained within CHILDES include data from both normal children and children with language disorders that span many different age groups. Importantly, CHILDES includes programs that make the database searchable for almost any possible research question, and is available free to the research community. A more thorough description of the CHILDES project can be found in MacWhinney (1996), and also at the CHILDES website, http://childes.psy.cmu.edu.

3.3.2 *British National Corpus* (*BNC*)

Originally released in 1995, the BNC is comprised of over 10 million words sampled from over 4,000 sources. Approximately 90% of the corpus is based on samples of written English taken from many different publication types and many different genres. Close to 1000 of the samples were taken from transcriptions of spoken English. As the name implies, all of the text and speech samples appearing in the BNC are from British English. Notably, the BNC is heavily tagged for part-of-speech, making it a useful tool for determining how individual words may be used in text and conversation. Additional information about the BNC, including how to acquire access to it, may be found at http://www.natcorp.ox.ac.uk. This corpus is frequently used in both cognitive linguistics and psycholinguistics as a result of the fact that, given its sheer size, it may be the most representative official corpus-based sample of English in current existence.

3.3.3 *American National Corpus* (*ANC*)

The ANC, although not yet completed, is designed to mirror the BNC, but with American English. The compilers of the ANC aim to provide over 100 million words of text, and intend to match the BNC in terms of the types of sources and genres included in the corpus. Moreover, the compilers aim to include samples of language from additional sources such as web pages and American music. Upon its completion, the ANC will probably be the most representative corpus-based sample of American English. Additional information about the ANC, along with information regarding a preliminary 11 million word first release version, can be found at http://americannationalcorpus.org.

3.3.4 *Brown corpus of standard American English* (*Kucera & Francis 1967*)

Originally released in 1964, the Brown Corpus was the first major machine-readable corpus, although it is available in both book and computer-based format. The corpus is comprised of over 1 million words of standard American English, extracted from 500 texts of approximately 2,000 words apiece. The text samples, all published in 1961, were compiled from texts across many different genres and publication types. The Brown corpus is tagged for part-of-speech, affording accessibility to numerous sources of information about word usage in standard American English. Although it is somewhat small and outdated (compared to the BNC or the ANC), it is free, easily accessible on the internet (just search Google for "Brown Corpus of Standard American English," or some variant thereof) or in many university libraries, and is still used and cited often.

3.3.5 *Penn Treebank* (*PTB*)

The Penn Treebank corpus is notably different from the other English corpora listed above. The Penn Treebank is a collection of approximately 2,500 stories taken from the Wall Street Journal over three years, along with a tagged version of the Brown Corpus and a sample of spoken English. All text contained within the PTB is part-of-speech tagged. However, the added value of this corpus is that a large subset of the text contained within the PTB has been annotated with syntactic information, especially in the form of predicate/argument structure. This syntactic annotation makes the PTB especially appealing to researchers interested in studying specific syntactic trends in a sample of naturally-occurring language. More information about the Penn Treebank project may be found at http://www.cis.upenn.edu/~treebank/home.html.

3.3.6 *An alternative to formal corpora*

Here, we offer a brief caveat to people intending to work with corpora in order to extract frequency information, along with, what we hope, is an adequate solution. Through analyzing frequency information in any of the corpora listed above, there are some (possibly many) cases whereby the frequency counts may be under-representative of realistic usage patterns. Consider *perfectly honest* versus *perfectly unfair*. A search of the BNC reveals that the bigram probability for each word-pair is zero. That is, according to the BNC, *perfectly* never appears directly before either *honest* or *unfair*. However, a search of Google reveals a very different story. As of February 21, 2004, *perfectly* appears immediately before *honest* approximately 151,000 times, whereas it appears directly before *unfair* 107 times. Thus, by searching Google, it becomes evident that although both *perfectly honest* and *perfectly unfair* are in use, there is a massive asymmetry in their usage patterns. This noted asymmetry could have posed a problem for any experimenter who wished to study these word-pairs, but assumed, from, say BNC frequency counts, that neither item appeared in naturally occurring English.

It should be noted that marked disparities between large English corpora and Google are probably most notable in situations where the words under investigation are very low in frequency, or when the researcher is interested in bigram probabilities of words that wouldn't be expected to co-occur regularly. Indeed, Zhu and Rosenfeld (2001) have noted that the occurrence of specific triples of words (trigrams) is quite rare, even in relatively large corpora. The existence of such disparities brings to light an interesting question, "What is the feasibility of using Google as an English corpus?" In other words, could it be possible, and more importantly, empirically justified, to use internet search engines as a way to overcome some of the data sparseness problems noted above? Although many members of the language science community may express some reticence, research demonstrating the efficacy of Google to produce valid and reliable frequency counts of English does exist.

Keller and Lapata (2003), for example, identified a set of bigrams and then determined their prevalence in the BNC and the North American News Text Corpus (NANTC). They subsequently conducted very simple internet searches of those same bigrams using both Google and Altavista search engines. They found that the bigram frequencies extracted from the two internet search engines correlate very highly with the bigram information extracted from the two formal corpora. More impressively, they found that the bigram frequency data extracted from the internet search engines correlate with human plausibility judgments even better than the bigram frequency information extracted from the two corpora. This evidence, considered together, can be taken as a compelling demonstration of the potential utility of the web for examining naturally occurring language patterns.

Internet searches are quick, simple, free, readily available, and, given the enormous number of sites on the internet, allow researchers to overcome data sparseness problems observed with corpora. For these reasons, and in light of recent empirical studies, we advocate the use of internet search engines for simple frequency counts. Internet search engines are not, however, without their problems. Most notably, and most unfortunately, the internet is not a tagged corpus. As such, information about parts-of-speech, specific linguistic features, and syntactic usage is not available. Additionally, English and Chinese appear to be the dominant languages of the internet; thus, the data sparseness problems alluded to above may not be overcome for searches in other languages. Nevertheless, internet search engines are a great way to quickly examine hypotheses regarding patterns of usage, and are a reasonable first-step in attempts to quantify current language use.

3.4 Summary

Given the multifarious manner in which corpora may be used, it is difficult to identify a precise set of steps by which one might go about analyzing a corpus. However, it is clear that initially, one should chose a corpus to analyze by considering both pragmatic and representativeness issues. Searching a corpus depends on the format of a chosen corpus, along with the existence of some sort of software, compatible with the corpus, to which the researcher can gain access. Luckily, many corpora available on the internet (or purchasable) also have search-command interfaces that make extracting the needed information from the corpus relatively easy. For most of the corpora listed above, such search-command interfaces exist, and are detailed in the links provided. Of course, if one were to use Google, the corpus is easily searched by regular search strategies familiar to people with a moderate level of contemporary computing knowledge.

4. Concluding remarks

Through the information presented within this chapter, it should be evident that examining 'What People Actually Say' can serve to enlighten many aspects of linguistic and psycholinguistic theory. Here, we have provided an introduction to the complexities involved in working with both authentic usage data and in examining existing corpora. The interdisciplinary body of research, based on authentic usage data, considered in this chapter supports numerous tenets of cognitive-functionalist linguistics in different ways. In addition to exploring selected dynamic approaches to language (see also Waugh et al. this volume), we also saw that sign language and gesture research both demonstrate how studying visuo-spatial actional modalities can afford insights into embodied cognition in general, and manifestations of iconicity, metaphor, and spatial concepts in particular (see also the chapter on sign language by Morford and Wilcox, this volume, and the chapters on co-speech gesture by Sweetser and Mittelberg respectively). Furthermore, studying usage patterns found in large corpora yields information regarding the structuring of linguistic systems, and also regarding how members of a speech community employ these (multimodal) resources in social interaction. Moreover, we have reviewed the ways in which data extracted from corpora may be used to generate testable hypotheses and facilitate greater experimental control (see also Grondelaers et al. this volume).

Throughout the chapter, we have identified what seems to us a trend in combining qualitative and quantitative methods of analyzing authentic language data, be they in the written or spoken modality. That is, usage patterns may be identified by searching large corpora for specific lexical items or grammatical structures. Complementing such initial data-driven approaches with fine-grained analyses of the linguistic and, depending on the method, also extra-linguistic context can then reveal emergent patterns of interaction and how these may align with cultural schemata. We see this methodological interaction as a noteworthy and invaluable strength of usage-based approaches.

From the viewpoints inherent within the different methodologies presented here, it can be seen that linguistic knowledge is an intricate part of human cognitive abilities. All the various empirical approaches described, taken together, demonstrate the dynamic nature of human cognitive processing, and lead us to a more comprehensive picture of the cognitive, embodied, situated, and social aspects of both linguistic structure and language use.

References

Aijmer, K. & Stenström, A. (Eds.). (2004). *Discourse patterns in spoken and written corpora*. Amsterdam/ Philadelphia: John Benjamins.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London, UK: Routledge.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam: John Benjamins.

Bakhtin, M. M. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.

Barodal, J. (2005). A crash course in corpus linguistics. Retrieved May 12, 2006 from www.ling.unt.edu/ corpus.html

Barlow, M. & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, CA: CSLI Publications.

Beaugarde, R. (1997). The story of discourse analysis. In T. van Dijk (Ed.), *Discourse as structure and process. Discourse studies: A multidisciplinary introduction* (pp. 35–62). London/Thousand Oaks: Sage.

Biber, D. (2003). Variation among university spoken and written registers: A new multi-dimensional analysis. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 47–70). Amsterdam/New York: Rodopi.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Botley, S. P., McEnery, A. M., & Wilson, A. (Eds.). (2000). *Multilingual corpora in teaching and research*. Atlanta, GA: Rodopi.

Bouvet, D. (1997). *Le corps et la métaphore dans les langues gestuelles: A la recherche des modes de production des signes*. Paris: L'Harmattan.

Bouvet, D. (2001). *La dimension corporelle de la parole. Les marques posturo-mimo-gestuelles de la parole, leurs aspects métonymiques et métaphoriques, et leur rôle au cours d'un récit.* Paris: Peeters.

Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 145–167). Mahwah, NJ: Erlbaum.

Bybee, J. & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure.* Amsterdam/ Philadelphia: John Benjamins.

Calbris, G. (2003). From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture, 3*, 19–46.

Cameron, L. & Low, G. (Eds.). (1999). *Researching and applying metaphor*. Cambridge: Cambridge University Press.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam/Philadelphia: John Benjamins.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: University of Chicago Press.

Chafe, W. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 93–111). Mahwah, NJ: Erlbaum.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cienki, A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J. P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 189–204). Stanford: CSLI Publications.

Cienki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), in cooperation with J. Grady, *From perception to meaning: Image schemas in cognitive linguistics* (pp. 421–442). Berlin/New York: Mouton de Gruyter.

Cienki, A., Luka, B. J., & Smith, M. B. (Eds.). (2001). *Conceptual and discourse factors in linguistic structure*. Stanford: Center for the Study of Language and Information.

Cienki, A. & Müller, C. (Eds.). (fc.). *Metaphor and gesture*. Amsterdam/Philadelphia: John Benjamins.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Coffin, C., Hewings, A., & O'Halloran, K. (Eds.). (2004). *Applying English grammar: Functional and corpus approaches*. London: Arnold Publishers.

Coulson, S. & Oakley, T. (2000). Blending basics. *Cognitive Linguistics*, *11*, 175–196.

Croft, W. & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Deignan, A. (1999). Corpus-based research into metaphor. In L. J. Cameron & G. D. Low (Eds.), *Researching and applying metaphor* (pp. 177–199). Cambridge: Cambridge University Press.

Dirven, R., Polzenhagen, F., & Wolf, H. (fc./a). Cognitive linguistics, ideology, and critical discourse analysis. In D. Geeraerts & H. Cuyckens (Eds.), *Handbook of cognitive linguistics*. Oxford: Oxford University Press.

Dirven, R., Wolf, H., & Polzenhagen, F. (fc./b). Cultural studies. In D. Geeraerts & H. Cuyckens (Eds.), *Handbook of Cognitive Linguistics*. Oxford: Oxford University Press.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language, 63*, 805–855.

Du Bois, J. W. (2000). *Santa Barbara corpus of spoken American English, Part 1.* Philadelphia: Linguistic Data Consortium.

Du Bois, J. W. (2003a). Discourse and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 47–87). Mahwah, NJ: Erlbaum.

Du Bois, J. W. (2003b). Argument structure: Grammar in use. In J. W. Du Bois, L. E. Kumpf, & W. J. Ashby (Eds.), *Preferred argument structure: Grammar as architecture for function* (pp. 11–60). Amsterdam/Philadelphia: John Benjamins.

Du Bois, J. W. (Ed.). (2003c). *Santa Barbara corpus of spoken American English*, *Part 2*. Philadelphia: Linguistic Data Consortium.

Du Bois, J. W., Kumpf, L. E., & Ashby, W. J. (Eds.). (2003). *Preferred argument structure: Grammar as architecture for function.* Amsterdam/Philadelphia: John Benjamins.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–87). Mahwah, NJ: Lawrence Erlbaum Assoc.

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, *15*, 223–238.

Duncan, S. (2003). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture, 2*, 183–206.

Duranti, A. (1997). *Linguistic anthropology.* Cambridge: Cambridge University Press.

Duranti, A. & Goodwin, C. (Eds.). (1992). *Rethinking context: Language as an interactive phenomenon.* Cambridge: Cambridge University Press.

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of the mind*. New York: HarperCollins.

Erickson, F. (1985). Audiovisual records as a primary data source. *Sociological Methods & Research*, *2*, 213–232.

Fairclough, N. L. (1992). *Discourse and social change.* Cambridge: Polity Press.

Fairclough, N. L. (1995a). *Critical discourse analysis: The critical study of language*. Harlow, UK: Longman.

Fairclough, N. L. (1995b). *Media discourse*. London: Edward Arnold.

Fairclough, N. L. & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as structure and process. Discourse studies: A multidisciplinary introduction* (pp. 258–284). London/Thousand Oaks/New Delhi: Sage Publications.

Fasold, R. (1990). *Sociolinguistics of language*. Cambridge, MA: Blackwell.

Fauconnier, G. (1994). *Mental spaces*. Cambridge, MA: Cambridge University Press.

Fauconnier, G. & Turner, M. (2002). *The way we think: Conceptual integration and the mind's hidden complexities*. New York, NY: Basis Books.

Fleischman, S. & Waugh, L. R. (Eds.). (1991). *Discourse pragmatics and the verb: Evidence from romance*. London/New York: Routledge.

Fonseca-Greber, B. & Waugh, L. R. (2003). On the radical difference between the subject personal pronouns in written and spoken european french. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 225–240). Amsterdam/New York: Rodopi.

Ford, C. E., Fox, B. A., & Thompson, S. A. (2003). Social interaction and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 119–143). Mahwah, NJ: Erlbaum.

Fox, B. A. (1994). Contextualization, indexicality, and the distributed nature of grammar. *Language Sciences*, *16*, 1–38.

Fox, B. A., Hayashi, M., & Jasperson, R. (1996). A cross-linguistic study of syntax and repair. In E. Ochs, E. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 185–237). Cambridge: Cambridge University Press.

Frazier, L. & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*, 505–526.

Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instructions. In D. McNeill (Ed.), *Language and gesture* (pp. 99–117). Cambridge: Cambridge University Press.

Ghadessy, M., Henry, A., & Roseberry, R. (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins.

Gibbs, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. New York, NY: Cambridge University Press.

Gibbs, R. W. (2003). Embodied experience and linguistic meaning. *Brain and Language, 84*, 1–15.

Giddens, A. (1984). *The constitution of society: Outline of a theory of structuration*. Cambridge, UK: Polity Press.

Givón, T. (Ed.). (1987). *Topic continuity in discourse: A quantitative cross-language study.* Amsterdam: John Benjamins.

Glasbey, S. R., Barnden, J. A., Lee, M. G., & Wallington, A. M. (2002). Temporal metaphors in discourse. Paper presented at the Conference on Metaphor and Thought, Pontifical Catholic University of Sao Paulo, Brazil.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). Switchboard: Telephone speech corpus for research and development. *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, *1*, 517–520.

Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Goldin-Meadow, S. (2003). *Hearing gesture: How hands help us think*. Cambridge, MA/London: The Belknap Press of Harvard University Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York, NY: Irvington.

Goodwin, C. (1981). *Conversational interaction*. New York, NY: Academic Press.

Goodwin, C. (2001). Practices of seeing visual analysis: An ethnomethodological approach. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of visual analysis* (pp. 157–182). London: Sage Publications.

Grady, J. & Johnson, C. (2002). Converging evidence for the notions of *subscene* and *primary scene*. In R. Dirven & R. Pörings (Eds.), *Metaphor and metonymy in comparison and contrast* (pp. 533–554). Berlin/New York: Mouton de Gruyter.

Grady, J., Oakley, T., & Coulson, S. (1999). Blending and metaphor. In G. Steen & R. Gibbs (Eds.), *Metaphor in cognitive linguistics* (pp. 101–124). Amsterdam/Philadelphia: John Benjamins.

Gries, S. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics, 1*, 1–27.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Gumperz, J. J. (2001). Interactional sociolinguistics: A personal perspective. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 215–228). Malden, MA/Oxford, UK: Blackwell.

Gumperz, J. J. & Berenz, N. (1993). Transcribing conversational exchanges. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 91–121). Hillsdale, NJ: Lawrence Erlbaum.

Hanks, W. F. (1996). *Language and communicative power*. Boulder, CO: Westview Press.

Hodge, R. & Kress, G. (1988). *Social semiotics*. Ithaca, NY: Cornell University Press.

Hougaard, A. (2005). How're we doing: An interactional approach to cognitive processes of online meaning construction. Unpublished doctoral dissertation, University of Southern Denmark, Odense.

Hopper, P. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–175). Mahwah, NJ: Lawrence Erlbaum.

Hopper, P. & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, *56*, 251–299.

Hughes, R., Carter, R., & McCarthy, M. (1995). Discourse context as a predictor of grammatical choice. In D. Graddol & T. Stephen (Eds.), *Language in a changing Europe* (pp. 47–54). Clevedon: British Assoc for Applied Linguistics.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hutchins, E. (2002). Material anchors for conceptual blends. In Proceedings of the Way We Think Conference on Conceptual Blending. University of Southern Denmark, Odense. *Odense Working Papers in Language and Communication, 23,* 85–111.

Hymes, D. (1961). The ethnography of speaking. In T. Gladwin & W. C. Sturtevant (Eds.), *Anthropology and human behavior*. Washington, DC: Anthropological Society of Washington, DC (Reprinted in J. A. Fishman (Ed.), 1968, *Readings in the Sociology of Language* (pp. 99–119). The Hague: Mouton.

Jaworski, A. & Coupland, N. (Eds.). (1999). *The discourse reader*. London/New York: Routledge.

Johnson, M. (1987). *The Body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago, IL: University of Chicago Press.

Johnstone, B. (2002). *Discourse analysis*. Oxford: Blackwell.

Keller, F. & Lapata, M. (2003). Using the web to obtain frequencies for unseen bigrams. *Computational Linguistics, 29*, 459–484.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters.* Cambridge/New York: Cambridge University Press.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversations. *Journal of Pragmatics, 23,* 247–279.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and Gesture* (pp. 47–63). Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London/New York: Longman.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Koenig, J. (Ed.). (1998). *Discourse and cognition: Bridging the gap*. Stanford, CA: CSLI Publications.

Kövecses, Z. (2002). *Metaphor: A practical introduction*. Oxford/New York: Oxford University Press.

Kövecses, Z. (2005). *Metaphor in culture: Universality and variation*. Cambridge, MA/New York: Cambridge University Press.

Kramsch, C. (2002). 'How can we tell the dancer form the dance?' In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 1–30). London/New York: Continuum.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London/New York: Continuum.

Kress, G. & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London/New York: Routledge.

Kucera, H. & Francis, W. N. (1967). *Computational analysis of present day American English.* Providence, RI: Brown University Press.

Labov, W. (1972). *Language in the inner city: Studies in the black English vernacular.* Philadelphia: University of Pennsylvania Press.

Lakoff, G. (1987). *Women, fire, and dangerous things. What categories reveal about the mind.* Chicago, IL: University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–251). Cambridge: Cambridge University Press.

Lakoff, R. T. (2001). Nine ways of looking at apologies: The necessity for interdisciplinary theory and method in discourse analysis. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 199–214). Malden, MA/Oxford, UK: Blackwell.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York, NY: Basic Books.

Lakoff, G. & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. Chicago, IL: Chicago University Press.

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford: Stanford University Press.

Langacker, R.W. (2001). Discourse in cognitive grammar. *Cognitive Linguistics*, *12*, 143–188.

Lantolf, J. & Thorne, S. (Eds.). (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

Leistyna, P. & Meyer, C. F. (Eds.). (2003). *Corpus analysis: Language structure and language use.* Amsterdam/New York: Rodopi.

Liddell, S. (1998). Grounded blends, gestures, and conceptual shifts. *Cognitive Linguistics*, *9,* 283–314.

Liddell, S. (2003). *Grammar, gesture, and meaning in American Sign Language.* Cambridge: Cambridge University Press.

Liebert, W., Redeker, G., & Waugh, L. R. (Eds.). (1997). *Discourse and perspective in cognitive linguistics.* Amsterdam/Philadelphia: John Benjamins.

MacDonald, M. C. (1993). The interaction of lexical and syntactic ambiguity. *Journal of Memory and Language, 32*, 692–715.

MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology, 5*, 5–14.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Manning, C. D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, MA: The MIT Press.

McEnery, T. & Wilson, A. (1996). *Corpus linguistics*. Edinburgh, UK: Edinburgh University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago, IL: Chicago University Press.

McNeill, D. (Ed.). (2000). *Language and gesture*. Cambridge: Cambridge University Press.

McNeill, D. (2003). Aspects of aspect. *Gesture*, *3*, 1–17.

McNeill, D. (2005). *Gesture and thought*. Chicago, IL: Chicago University Press.

Meyer, C. F. (2002). *English corpus linguistics. An introduction*. Cambridge: Cambridge University Press.

Mittelberg, I. (2006). Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar. Doctoral Dissertation, Cornell University.

Mittelberg, I. (fc.). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture*. Amsterdam/Philadelphia: John Benjamins.

Moerman, M. (1988). *Talking culture: Ethnography and conversation analysis*. Philadelphia, PA: University of Philadelphia Press.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich.* Berlin: Berlin Verlag A. Spitz.

Müller, C. (2004). Forms and uses of the palm up open hand: A case of a gesture family? In R. Posner & C. Müller (Eds.), *The semantics and pragmatics of everyday gestures. The Berlin conference* (pp. 233–256). Berlin: Weidler Verlag.

Musolff, A. (2004). *Metaphor and political discourse. Analogical reasoning in debates about Europe*. Basingstoke, Hampshire/New York: Palgrave McMillan.

Nerlich, B., Hamilton, C. A., & Rowe, V. (2002). Conceptualising foot and mouth disease: The socio-cultural role of metaphors, frames and narratives. *Metaphorik.de, 2*, 90–108.

Núñez, R. E. (2004). Do real numbers really move? Language, thought, and gesture: The embodied cognitive foundations of mathematics. In F. Iida, R. Pfeifer, L. Steels, & Y. Kuniyoshi (Eds.), *Embodied artificial intelligence* (pp. 54–73). Berlin: Springer-Verlag.

Núñez, R. E. & Sweetser, E. E. (2006). Aymara, where the future is behind you: Convergent evidence from language and gesture in the cross-linguistic comparison of spatial construals of time. *Cognitive Science, 30*, 1–49.

Ochs, E., Schegloff, E. A., & Thompson, S. A. (Eds.). (1996). *Interaction and grammar*. Cambridge: Cambridge University Press.

Parrill, F. & Sweetser, E. E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, *4*, 197–219.

Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam/New York: John Benjamins.

Pascual, E. (2002). *Imaginary trialogues: Conceptual blending and fictive interaction in criminal courts*. Utrecht, The Netherlands: Landelijke Onderzoekschool Taalwetenschap.

Renkema, J. (2004). *Introduction to discourse studies*. Amsterdam: John Benjamins.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*, 696–735.

Schegloff, E. A. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social Psychology Quarterly*, *50*, 101–114.

Schegloff, E. A. (1993). Reflections on language, development, and the interactional character of talk-in-interaction. *Research on Language and Social Interaction*, *26*, 139–153.

Schegloff, E. A. (1996). Turn organization: One direction for inquiry into grammar and interaction. In E. Ochs, E. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 52–133). Cambridge: Cambridge University Press.

Schegloff, E. A. (2001). Discourse as interactional achievement III: The omnirelevance of action. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 229–249). Malden, MA/Oxford, UK: Blackwell.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D. (1994). *Approaches to discourse*. Oxford: Blackwell.

Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds.). (2001). *The handbook of discourse analysis*. Malden, MA/Oxford, UK: Blackwell.

Schuetze-Coburn, S. (1987). *Topic management and the lexicon: A discourse profile of three-argument verbs in German*. Unpublished M.A. thesis, Los Angeles: UCLA.

Settekorn, W. (2001a). Tor des Monats – Tor zur Welt. Zum Metapherngebrauch in Massenmedien. In D. Möhn, D. Ross, & M. Tjarkssobhani (Eds.), *Mediensprache und Medienlinguistik. Festschrift für Jörg Hennig* (pp. 93–109). Frankfurt/M.: Peter Lang.

Settekorn, W. (2001b). Media and metaphors: The case of virtual wandering and stationary movement. *Metaphorik.de, 1*, 87–99.

Settekorn, W. (2003). Representation: The influence of history, culture, and (mass) media on the evolution of a concept central to cognitive science. Presentation given at EMCL Workshop, Cornell University.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinha, C. (1999). Grounding, mapping, and acts of meaning. In T. Janssen & G. Redeker (Eds.), *Cognitive linguistics: Foundations, scope, and methodology* (pp. 223–255). Berlin: Mouton de Gruyter.

Slobin, D. (1996). From 'thought and language' to 'thinking for speaking.' In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 97–114). Cambridge: Cambridge University Press.

Smith, N. (2003). *Gesture and beyond.* Honors Thesis, Program in Cognitive Science, University of California at Berkeley.

Steen, G. (1994). *Understanding metaphor in literature: An empirical approach*. London/New York: Longman.

Steen, G. (1999). From linguistic to conceptual metaphor in five steps. In R. W. Gibbs & G. Steen (Eds.), *Metaphor in cognitive linguistics* (pp. 57–77). Amsterdam/Philadelphia: John Benjamins.

Steen, G. (2004a). Perspectives on discourse: State of the art. *Language and Literature*, *13*, 161–179.

Steen, G. (2004b). Can discourse properties of metaphor affect metaphor recognition? *Journal of Pragmatics*, *26*, 1295–1313.

Stefanowitsch, A. & Gries, S. T. (Eds.). (2006). *Corpus-based approaches to metaphor and metonymy.* Berlin/New York: Mouton de Gruyter.

Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs*, *60*, 275–299.

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction*, *3*, 239–276.

Stubbs, M. (1983). *Discourse analysis: The sociolinguistic analysis of natural language.* Chicago, IL: University of Chicago Press.

Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Oxford, UK/Cambridge/MA: Blackwell.

Sweetser, E. E. (1990). From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure. Cambridge: Cambridge University Press.

Sweetser, E. E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16e Congrès International des Linguistes* (CD-ROM), Elsevier.

Tabensky, A. (2001). Gesture and speech rephrasing in conversation. *Gesture*, *1*, 213–235.

Tannen, D. (1990). *Talking voices: Repetition, dialogue, and imagery in conversational discourse.* Cambridge: Cambridge University Press.

Tao, H. (1996). *Units in mandarin conversation: Prosody, discourse, and grammar.* Amsterdam/Philadelphia: John Benjamins.

Tao, H. (2003a). A usage-based approach to argument structure: 'remember' and 'forget' in spoken English. *International Journal of Corpus Linguistics*, *8*, 75–95.

Tao, H. (2003b). Turn initiators in spoken English: A corpus-based approach to interaction and grammar. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 187–207). Amsterdam/New York: Rodopi.

Taub, S. (2001). Language from the body: Iconicity and metaphor in American Sign Language. Cambridge: Cambridge University Press.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam/ Philadelphia: John Benjamins.

Tomasello, M. (Ed.). (1998, 2003). *The new psychology of language: Cognitive and functional approaches to language structure.* Mahwah, NJ: Erlbaum.

Turner, M. (1987). *Death is the mother of beauty: Mind, metaphor, criticism*. Chicago, IL: Chicago University Press.

van Dijk, T. A. (Ed.). (1997a). *Discourse as structure and process. Discourse studies: A multidisciplinary introduction.* Volume 1. London/Thousand Oaks: Sage.

van Dijk, T. A. (Ed.). (1997b). *Discourse as social interaction. Discourse studies: A multidisciplinary introduction.* Volume 2. London/Thousand Oaks: Sage.

van Dijk, T. A. (2001). Critical discourse analysis. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 352–371). Malden, MA/Oxford, UK: Blackwell.

van Hoek, K., Kibrik, A. A., & Nordman, L. (Eds.). (1999). *Discourse studies in cognitive linguistics. Selected Papers from the fifth International Cognitive Linguistics Conference, Amsterdam, July 1997.* Amsterdam/ Philadelphia: John Benjamins.

Wallington, A. M., Barnden, J. A., Buchlovsky, P., Fellows, L., & Glasbey, S. R. (2004). Metaphor annotation: A systematic study. (Technical Report CSRP-03-04).

Waugh, L. R. (1995). Reported speech in journalistic discourse: The relation of function and text. *Text, 15*, 129–173.

Wetherell, M., Taylor, S., & Yates, S. J. (Eds.). (2001). *Discourse as data: A guide for analysis*. London/ Thousand Oaks/New Delhi: Sage.

Wilcox, P. P. (2000). *Metaphor in American Sign Language*. Washington, DC: Gallaudet University Press.

Wilcox, P. P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics, 15*, 197–222.

Wilcox, S. (2004). Cognitive iconicity: Conceptual spaces, meaning, and gesture in signed languages. *Cognitive Linguistics*, *15*, 119–147.

Williams, R. (2004). *Making meaning from a clock: Material artifacts and conceptual blending in time-telling instruction*. Unpublished doctoral dissertation, University of California at San Diego.

Wilson, A., Rayson, P., & McEnery, T. (Eds.). (2003). *A rainbow of corpora: Corpus linguistics and the languages of the world*. München: Lincom.

Zinken, J. (2004). Metaphoric practices in the German *Wende* discourse. *Journal of Multilingual and Multicultural Development*, *25*, 1–13.

Zinken, J., Hellsten, I., & Nerlich, B. (fc.). Discourse metaphors. In R. Dirven, R. Frank, T. Ziemke, & J. Zlatev (Eds.), *Body, language, and mind. Vol. 2: Social situatedness*. Berlin: Mouton de Gruyter.

Zhu, X. & Rosenfeld, R. (2001). Improving trigram language modeling with the world wide web. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. Salt Lake City, Utah.

An introduction to experimental methods for language researchers

Monica Gonzalez-Marquez, Raymond B. Becker
and James E. Cutting

1. Introduction

Imagine yourself as a field researcher sent to Africa to study lion colonies. The savannah can be treacherous in its monotony. Inching forward through the tall grasses you suddenly discover yourself face to face with the snarling leader of a lion tribe. Would you say, "Pardon me, but I would like to know a little more about how you communicate with the other members of your family. For starters, how is it that when you, say, roar, that comes to mean 'an elephant is coming'?" Maybe you would have a flashback to your philosophical coursework, and begin to wonder about Wittgenstein's famous aphorism, "If a lion could talk, we could not understand him." Clearly, *asking* a lion how he communicates would be fruitless. What you would need to do is arrange controlled circumstances under which you could record his vocalizations in response to situations you construct and whose meaning you do understand. For somewhat similar reasons, there are serious epistemic weaknesses to simply asking a *human being* how she communicates. Although she may be able to reply in a language that you understand, the introspections that she shares with you will have undergone a great deal of perceptual, cognitive, and social alterations and filtrations. Her subjective report of how she uses language will only be a shadow of how her mind really uses it. To make matters worst, we know that people are not good at understanding why it is they behave in certain ways (Wilson & Nisbett). For cognitive linguists, this issue has particular implications, specifically those of embodiment theory. The following sections will flesh out how to go about asking similar questions, albeit about people, within the confines of the scientific method. They will also address practical concerns that should serve as a preliminary guide to doing experimental work.

2. Experimenting with experimentation

How do you do an experiment? How do you figure out what to study, which questions to ask, which methods to use to ask them? Answering these and all of the other related questions adequately, takes many years of training, false starts and successes. It all begins however, with learning to read a research paper. The approach can be thought of rather like reverse engineering, as each section of a research paper has a relationship to the research process. In the following sections we will deconstruct the research article with two goals in mind, to help you, as a new consumer of research literature, better understand published studies, and to help you, as a new experimentalist, design and develop your own experiments. As such, the material will first be presented for the benefit of a naïve reader followed by elaboration for the experimentalist. In the first section we recommend that you concentrate on the structure of the paper. Do not worry if you cannot follow the arguments well. What is important is to become familiar with the way a research story is told. Anything else that you manage to grasp will be icing on the cake. In the second section try to put yourself in the shoes of the researcher conducting the project by trying to understand why the researcher made the choices she did. Your ultimate goal, after all, is to understand the nature of those choices well enough to be able to conduct your own experiments. All of the information provided here should be general enough to apply to any type of behavioral research. To benefit most from this chapter, we recommend that you follow with a copy of a published experimental research article. We recommend those from a prestigious journal such as "Psychological Science" or "Cognitive Science."[1]

3. Reading the research article

"Stephen King's new novel will be available this Saturday at Modern Times Bookstore." This announcement may sound appealing to a Stephen King fan or inconsequential to someone who is not. Intimidating, overwhelming and confusing are not words usually evoked at the thought of actually reading the book. Many people would use precisely these terms, however, to describing reading a research article. There are reasons for this. Most people know what to expect from a novel. A novel is the telling of a story, and since King writes popular fiction, it is expected that the plot will contain a given set of elements that will be linked to each other in predictable ways. What makes a research article so intimidating to some readers is that they do not know what to expect. The American Psychological Association publishes a handbook that describes the general structure of articles. Though the handbook is aimed at researchers writing up their findings, following is an adaptation intended to help new consumers of psychological research know what to expect from a scientific research article.

The analogy between reading a novel and reading a research article was intended to help you conceptualize an article as a story unfolding paragraph by paragraph. As in popular fiction, when we read a paper, what we want to know is, 1. What is the setting, 2. who did what to whom, where, why, how and with what, 3. what was the outcome, and 4. what

1. Those readers who do not have easy access to scientific journals will find that the proceedings of the Cognitive Science Society conference are available free online at http://www.cognitivesciencesociety.org/cogsci.html

does the outcome mean for what we thought the story was. This information can usually be found in predictable places. Research articles, unlike (unfortunately!) most novels, have very strict space limitations. This means that every piece of information is significant. Everything from the title to the list of references will be relevant to understanding the piece of research as a whole.

3.1 Title

The title of an article serves two purposes, to inform the reader as briefly as possible about the subject matter of the paper, and to facilitate indexing in databases. Because authors have a very limited amount of space, they tend to use key terms in an eye-catching kind of way. For example, if a researcher were writing up an investigation into how bilingual children respond to image schemas while listening to stories in both the languages they speak, it is very likely that the title might be something like "Image Schemas and Narrative in Bilingual Children." This title makes it amply clear to readers interested in any of these three topics that this paper might be relevant to their needs. In addition some journals now allow inputting keywords for indexing.

3.2 Authors and affiliations

Authors listed are responsible for the research reported. The order of the authors reflects the agreed upon level of responsibility for the published work. The first author is most responsible and so forth. Affiliations typically reflect the institution that is the primary residence of each author. These might be academic institutions such as Cornell University, or they can be private research facilities such as the Santa Fe Institute. In addition, clinicians and others without institutional affiliation simply list their home city.

3.3 Place of publication

Different journals publish different types of research. Some are also more prestigious than others. Prestige is determined by different factors though the most important is consistent high quality research based on high selectivity. Examples of well respected journals that will be of interest to cognitive scientists in general include *Psychological Science*, *Trends in Cognitive Science*, and the various *Journal of Experimental Psychology* publications. As you read more research articles, it will become clear which journals are most relevant to your needs.

3.4 Year of publication

The year of publication is important to understanding general trends in research over time. Both within the context of an individual researcher's work as well as compared to other research, it will help you track the order in which findings emerged so as to better conceptualize the grounding of the findings.

As you become better acquainted with the literature of a certain field, it will become important for you to be able to reference different studies. These 4 pieces of information typically serve this purpose. It is considered good practice to commit to memory the authors, and the year, the publication outlet, and the title of a paper in decreasing order of importance. Google (either the general search engine or the specialized Google Scholar) will help you find the paper generally with the first and second pieces of information, and easily by including the third. Having this information readily available to memory will prove invaluable in activities ranging from preparing a study to having a conversation with a colleague on a research topic.

3.5 Abstract

An abstract at about 120 to 150 words is a concise summary of a research article. Though dense, it is generally well organized and is expected to contain specific types of information. The reader can expect to find what the main problem under investigation is, sometimes the characteristics of the subjects in the experiments, including sex and language abilities if relevant, etc., a description of the experimental method used, the findings, and the conclusions and implications. Like the title, the content of the abstract is also used for indexing in databases.

3.6 Introduction

The introduction is the first step into the paper itself. It is meant to set the stage for the rest of the paper by providing the reader with three pieces of information.

1. What the research question is.
2. What the literature has to say about it.
3. What the authors have done and why.

The research statement will be described at the beginning, usually very briefly. The literature review will involve presenting brief summaries of key studies related to the research question. Care will usually be taken to only discuss studies that are directly relevant. In addition, only the relevant parts of the studies themselves will be discussed. This point is especially important to keep in mind if you wish to do research along similar lines since it means that in order to get a complete picture of a cited article you will have to read the original article and should not rely on a brief summary. The final part of the introduction will give the authors' hypotheses and their motivation for conducting the research. In other words, they will briefly describe why they think the research is necessary.

3.7 Methods

The methods section is in many ways the 'proof is in the pudding' section in that it gives enough information about the experiments conducted so that anyone else can also conduct them (more about why this is important below). Details about the experiments will fall into these categories:

1. subjects
2. apparatus
3. procedure

3.7.1 *Subjects*

The subjects section will describe the subjects' characteristics. Typically demographics will be given along with the number of subjects, and the method used to recruit them. In addition, authors are encouraged to include information such as sex/gender. ethnic/cultural/national origin, socio-economic status (SES), and linguistic abilities. The rationale is that though these details may not be the target variables in the current studies, they might be relevant to other researchers for future studies. For example, a paper might say "50 Spanish/English early bilinguals participated for course credit. Their average age was 20 years of age with a range between 18 and 26 years. 25 were male and 25 were female. They were recruited from introductory psychology courses at the University of Texas at Austin. All subjects were of Mexican origin. Their SES was working class." Someone conducting a study five years hence might be investigating the impact of socio-economic status on cognition. The fact that the researchers above reported this information will prove valuable to the later researcher. Another important bit of information found in the procedure section is attrition rate. Authors will usually explain why any subjects were excluded, and how the criteria were set.

3.7.2 *Apparatus*

The apparatus or equipment used in an experiment can vary from the quite sophisticated such as a functional magnetic resonance imaging machine (a machine that uses blood flow to take pictures of activity in the brain) to the mundane such as playing cards. Regardless of what is used, all details necessary to obtaining the same tools will be provided. This includes model numbers of equipment as well as the names and versions of specialty software. If the equipment is particularly complex or unusual, an appendix may be added containing the necessary elaborated descriptions.

3.7.3 *Procedure*

The procedure section contains the details of exactly how and where the experiment took place. The grouping of the subjects, the nature of the stimuli, and the general breakdown of how the trials varied step by step is stated here. Figures, which are helpful for adding context to the description, can be found here as well. In brief, this section covers the steps from the beginning instructions to the debriefing.

3.8 Results

The results section provides a summary of the data collected using both descriptive and inferential statistics (see Nuñez this volume for an elaboration). The results will be given first in plain English and then in their statistical form as means, p-values, F-tests, etc. All analyses deemed necessary to support arguments in the discussion section will be presented here. In addition, any tables and graphs necessary to explain the results will also be found here.

3.9 Discussion

The previous parts of the article set all of the cards on the table, so to speak. It is in the discussion section that the authors will evaluate their results in relation to their hypotheses. This last point is particularly important because much of the disagreement about papers is specifically about whether the conclusions drawn follow from the data (or whether they can be explained with alternative explanations). In general, they will address points dealing with how their research was useful in resolving the target problem as well as any implications or consequences to the theory they are working in. Although this is the only place where anything resembling an opinion on the part of the authors is appropriate, most researchers will tend to be cautious about their claims, instead adopting the more productive position of encouraging further inquiry.

Many articles are composed of more than one experiment. When that is the case, you should expect short preliminary discussions to follow each experiment, with a general conclusion at the end.

3.10 References

The references section will list, in APA format, all sources cited in the research article. Aside from giving credit where credit is due, it is also an excellent starting point for further readings.

4. Other types of articles

The previous sections described the construction of an article reporting original research. There are two other main types of articles that you will encounter. The first is a review article and the second is a theoretical article. The review article is in many ways a tutorial of the subject matter. The author will review the available literature on a topic so as to evaluate its progress. You should expect to find summaries of key articles relating to the target question. A good review will also point out inconsistencies, contradictions, and gaps in the research, as well as recommendations for future research. They are especially useful for anyone considering research in a new area as well as for students rounding out their training.

A theoretical article shares many of the characteristics of a review article. The main difference is that where review articles tend to be geared toward seeing how well a research question has been addressed by the scientific community, a theoretical article will seek to support a theoretical model using extant research. Though most journals will publish both review and theoretical articles, especially good sources are *Behavioral and Brain Sciences*, *Perspectives on Psychological Science*, and especially *Psychological Review* for review articles and *Psychological Bulletin and Review* for theoretical articles. For beginning researchers, *Current Directions in Psychological Science* will prove a particularly accessible publication.

At this point you should be quite familiar with the type of information contained in the different parts of a research article. Hopefully you will have read not one, but several papers using the guidelines set out above and your mind is filled with questions about how and why things were done the way they were. Later, we will present additional criteria for reading journal articles that should help you develop solid experiments. Before continuing with the next section, following are a few preliminary details that should help you better conceptualize the nature of the experimental process.

5. The scientific method

> The purpose of the scientific method is to make sure nature hasn't misled you into thinking you know something that you don't actually know.
>
> Robert M. Pirsig. Zen and the Art of Motorcycle Maintenance, 1974

The word 'science' comes from the Latin *scire* meaning 'to know.' Science presents one of many ways of knowing. Other examples include the method of tenacity, the method of authority and a priori method (Rosnow & Rosenthal 1993). The method of tenacity refers to the manner in which people will cling to an idea simply because it seems to be common sense. The method of authority refers to taking what an authority figure such as a priest or a teacher says as fact without question. The a priori method refers to the use of pure reason and logic to come to conclusions about the world. Though all three are commonly used, they share a characteristic. The only verification they require is that of the individual having the belief. What sets science apart is its dependence on intersubjective verification, the possibility that knowledge can be empirically tested by different researchers. The Scientific Method, in general terms, is an intellectual framework geared at generating the most reliable findings possible as well as at facilitating their verification.

There are many misconceptions about how the scientific method is actually used. A common one is that it describes the step-by-step procedures followed unvaryingly by all scientific disciplines. The reality is quite different. Researchers use different methods developed specifically to understand their chunk of the environment (Tattersall 2002; Solso & Johnson 1984; Rosnow & Rosenthal 1993). The procedures used to investigate lava flow do not necessarily apply to understanding how neurons communicate nor to understanding how to cure cancer. They are all scientific methods insofar as their motivation is to produce reliable and verifiable findings, though it is doubtful that they share too many other practical characteristics. It is actually more appropriate to refer to the scientific method as scientific methodology (Solso & Johnson 1984) so as to make amply clear that there is actually a compendium of methods, all with a common goal.

Two other common misconceptions are actually different sides of the same coin. In the general description of scientific methodology, the possibility that findings could be verified by different researchers was given as an important component. The consensus possible from multiple testing has sometimes been misinterpreted as agreement by consensus. This type of agreement is also known as majority rule. Actually, since science is not a democracy, what it refers to is agreement by repeated recognition of low probability of falsehood (Tattersall 2002). Statistics is a central factor in reaching consensus (see Nuñez this volume). The general form that an accepted idea takes (and which will be further developed below) is that it has been shown to be very unlikely the result of chance. Specifically, that when different researchers went through the process of collecting data in comparable ways, and analyzed their results, the statistical analyses they conducted indicated a very low probability that the results were the product of random variation. Thus, there is consensus about the high probability that the phenomenon described by the results is not illusory, leading directly to the other side of the coin for this issue.

If you were to ask any person off the street what they thought was the goal of science, most would unfortunately respond that it was the pursuit of The Truth about the world. Not quite. Most scientists, in fact, believe that there is no way we will ever know the true nature of the world, a perspective formalized in the 20th century as a philosophical movement called Positivism. And that therefore, the best we can do is to develop possible explanations supported by empirical findings. The goal of science is then to find the best available explanation for a phenomenon, with 'best' qualified as being the least likely to be false. This is not the same as saying that it is most likely the correct explanation, but only that it is better than any other under consideration. It also means that it too can be dislodged the moment another explanation is found that better accounts for the target phenomenon.

Science as a human endeavor, prone to all of the subjectivities of the human experience, continues to be the topic of much debate. Though some would disagree, we propose that science is our attempt as a species to approach the closest thing to objectivity that we can achieve as subjective beings. Since our goal is to introduce the issues to the novice, we will limit discussion to just these few points. Readers interested in further elaboration are encouraged to consult Carnap's (1966) "The philosophy of science," Chalmers' (1976) "What is this thing called science?" Kuhn's (1962) "The structure of scientific revolutions," Lakatos' (1970) "Falsification and the methodology of scientific research programs," and Popper's (1959) 'The logic of scientific discovery."

6. Conducting research

Aronson, E., Wilson, T. W., & Brewer, M. B. (1998) claimed that there was no better way to understand the research process than by the careful reading of research articles. At the beginning of the chapter, you were introduced to the structure of a paper and were instructed in your initial readings to only pay attention to the general context and not necessarily the content of what was written. Understanding the research process via journal articles obviously requires more than structure based readings. Besides knowing where to look for information, it is also necessary to have the criteria to evaluate the contents. The basis for which depends on background knowledge of the subject matter as well as on an understanding of the nature of the different components involved in doing research. The first element can only be acquired through the deep study of the choice subject, the methodological aspects of which will hopefully be facilitated by the contents of this volume. The latter, along with a short repertoire of questions formulated to facilitate evaluation of each research component, is discussed below.

The following sections describe scientific methodology as developed to investigate cognitive phenomena. As above, the different steps will be presented within the structural context of the research article. In other words, each step will be discussed in reference to the section of a research article where it appears. This does not mean that there is a one-to-one correspondence between research steps and the sections of a paper, only that all the steps are represented in some form throughout. Our goal is to help you recognize them so that you will better understand the research process. (Note: All papers described in this section marked with "CS" are available for download at http://www.cognitivesciencesociety.org/cogsci.html. They can be found in the 2003, and 2004 proceedings.)

Conducting research generally involves the following steps, roughly divided in 2 parts (boldface indicates where the information can generally be found in a research article.)

Part One:
Introduction –

1. Narrowing down the topic of interest
2. Conducting an exhaustive literature review
3. Deciding on a question
4. Formulating a hypothesis

Part Two

1. **Methods** – developing an experiment
2. **Results** – analyzing data
3. **Discussion** – interpreting results

6.1 Part one: The introduction – the thinking and reading steps

The first 4 steps of conducting research are typically addressed in the introduction. They involve the project's conceptualization and motivation. Though they are given as steps, they might be better described as components in that their relationship with each other is interdependent. The ultimate goal is to produce a question that can be addressed experimentally. The decisions made along the way determine the general architecture of the research, and as such, should be considered carefully.

6.1.1 *Narrowing down the topic of interest*

The first step in conducting a research project is to narrow down the topic of interest. If you are reading this book, chances are very high that you have at least a couple of ideas that you would like to test experimentally. Chances are also very high that your ideas are far too broad to be addressed by one project, i.e. "what is the relationship between the body and language;" 'how is spatial information used by language," etc. This is common to new researchers. In fact, a large part of doing good experimental research involves learning how to ask questions that can be addressed experimentally. A good way to think of this process is as an inverted pyramid, the first step of which is determining your general area of interest, followed by deciding on a particular question and ending with the development of a specific hypothesis. Regardless, deciding which general phenomenon you want to study is the first step. An important consideration as a beginning researcher is that your topic have a history of being studied by other more experienced researchers. That is not to say that you should always shy away from topics that have not been investigated experimentally, only that in learning how to do research, working from well established models will prove infinitely useful. In short, you want to do as the experts have done.

Deciding what area you want to study begins with reading broadly. If you are interested in language and space, the first step is to introduce yourself generally to the issues. A good place to begin if you have no background might be with an undergraduate textbook. Sternberg's (2003), is a relatively up to date survey of cognitive psychology. The section on knowledge representation with its treatment of spatial cognition will prove quite enlightening if you are interested in the meaning structures underlying language. (Note to readers with a background in a different discipline: as you read the material, it will be very useful to suspend judgment about the perspective taken. You may not agree with the topics of study nor with the conclusions. Regardless, it will be key that you try to understand the perspective that led to the types of questions asked, a skill that will be pivotal to accurate critical assessment of all the research you encounter.) The next step might be more specialized books such as Bloom, Peterson, Nadel, and Garret's (1996) *Language and Space*, and Newcombe and Huttenlocher's (2000) *Making Space*. Books might be more appropriate as they tend to give more conceptual descriptions of research, and may be easier to digest for the research novice. Alternately, review articles by journals such as *Current Trends in Psychological Science* and *Psychonomics Bulletin & Review* will be useful. The goal is to find something that piques your curiosity.

6.1.2 *Conducting a literature review*

After deciding on a general topic, the next step is finding out what other researchers have already done, and that requires learning how to search. If you did begin your inquiry by reading a book or a review article, the bibliographies will be a very useful place to begin your background reading. Otherwise, the place to begin will be at the library using databases. Most libraries have online databases you can access either on site or via the internet. Each will have specific guidelines and tutorials for how to conduct searches. Walk yourself through these before beginning; it will save you much aggravation. Searching databases in general involves using keywords. Though your library will have guidelines to help you along, a place to start will be with target terms used in what you have read as general background in a textbook, etc. An alternate source involves using a cataloguing software such as Devonthink, which, among its many features, can provide you with frequency lists of the words used in an article. These often prove quite useful when conducting a search.

Besides library databases, other resources include http://scholar.google.com, a new search engine dedicated specifically to academic concerns. Doing a generalized 'google' search (http://www.google.com) can also be a good idea. In choosing what to read, however, avoid anything that has not been published in a reputable journal. Make sure, at the very least, that the piece was written by someone whose work has been referenced elsewhere or that you know.

Your goals in actually sitting down and reading the masses of papers your search will have produced are 1) to find out what has been done, 2) why it was done and 3) how it was done. Going section by section, following are questions you should ask yourself as you read through the articles (adapted from Jordan & Zanna 1999). A last note before beginning to read, it is probably not a good idea to read the sections of articles in sequential order. Instead start off by reading the abstract then the introduction followed by the discussion/conclusion. When the answers to questions 1 and 2 are reasonably clear, continue with the rest of the paper. If the paper has multiple experiments, skip to the general discussion at the end of the paper, then read the inner sections

6.1.3 *Questions to help you understand a research article*

6.1.3.1 *Abstract*

– What is the paper about?
– How did the researchers set up their study? i.e., who were the subjects, what was the experiment?
– What did the experiment measure?
– What were the main results of the study?

6.1.3.2 *Introduction*

– What were the theoretical considerations underlying the research?
– Why was the particular topic chosen for study?
– Does the chosen topic have implications beyond itself?
– What are the authors hypotheses?
– What questions do the researchers hope to answer with the results of their study? (Note that this is a different question than what their hypotheses were.)
– How did the authors decide on their research strategy, i.e. did they develop an experiment or chose to do a correlational study?

6.1.3.3 *Method*

– How were the hypotheses turned into testable questions?
– How were the variables manipulated, i.e. how was the experiment done?
– Were appropriate controls used?
– Were the measures used appropriate to the question being asked, i.e. is income an acceptable measure of socio-economic status?

6.1.3.4 *Results*

– What are the main results of the study?
– Can the results be used to answer the research question?
– Can the results be generalized beyond the context of the study?

6.1.3.5 *Discussion*

- What conclusions do the researchers draw from their results?
- What questions were left unanswered by the study?

After finishing the paper, should you have found it particularly insightful and interesting, you might want to do a background search on the authors to find out what else they have published. Finding out where their work has been cited will also be helpful in developing a clearer picture of how the research fits the broader scope of work on your topic and on cognition in general. A database such as the Web of Science (found at your library's online database directory) as well as Google Scholar can provide you with this information.

6.1.4 *Developing a research question*

Reviewing the literature is closely tied with choosing a research question. You can think of it rather like spirals that feed into each other. You begin with a vague idea of what you want to study, i.e. language and space, then you read a few papers. It strikes you that most of what has been written has been on the English language, so you decide you would like to look at prepositions in French. You go and find everything you can, theoretically and experimentally, on French prepositions. After reading even more, you decide you are interested in locative prepositions, etc. By the same token, as you fine tune your question, persist with your literature search. A researcher who uses experimentation is first a researcher. This means that your first goal is to find the answer to your question. It is possible, if not likely, that in reviewing the literature you will find the answer to your original question, in which case you will have to decide whether to extend the findings somehow or to change topic.

Once you have done at least a preliminary literature review, the next step is deciding what the question itself should be. What do you want to know? Be as precise as possible. There are issues to keep in mind as you concretize your thoughts. The question you decide upon should be one you find deeply interesting. Doing research takes a lot of time and patience. A deep interest will be required to keep you motivated through the rough spots. The question should also be relevant. You should carefully think through why the research is timely and important. A good question will always go beyond itself to help explain a larger phenomenon. Though it will focus on a seemingly small problem, it should contribute to our understanding of language and cognition. Ideally, your question should also be novel, elucidating a previously unexplored issue. If it cannot be completely novel, at the very least, it should shed further, necessary, light on a well-established topic.

There are also practical considerations. The question must be testable. There is a famous Sidney Harris cartoon showing two bespectacled, lab-coat-arrayed scientists discussing their new elixir. One says to the other, "It may well bring about eternal life, but it will take forever to test." Time is not the only obstacle to testability, of course. Questions such as 'Can consciousness be transferred from one person to another?' may be deeply interesting but, to the authors' knowledge, are currently far beyond our means to explore.

Given that your question is testable, research questions vary in the scope, or the level, at which the investigation will address the problem. The scope of the research question plays a balancing act with the practicality of finding an answer. As the scope gets wider, the more difficult it is to find a clear answer to a question. However, if the scope is too narrow, then the answer, although more easily found, might not be interesting.

Adequately addressing these issues will greatly facilitate the rest of the research process. A lack of clarity here will unavoidably be reflected in the finished product, i.e. the manuscript. You should only proceed with the more practical aspects of research after careful examination of the 'thinking' steps.

Ember and Ember (2001) lay out how to come up with a question that a researcher will be satisfied pursuing. Following are four types of research questions, using possible examples from cognitive linguistics.

> *1. Descriptive questions* are concerned with the prevalence of a particular phenomenon, i.e., What type of metaphors do people use to refer to temporal events in English versus Aymara? How is snow described in French as compared to Inuit?
>
> *2. Causal questions* are *why* types of questions, where the investigator wants to know the reason for the difference in languages, i.e., Why do Spanish speakers conflate manner of motion in the verb (e.g., *subir*), whereas English speakers use a satellite-frame construction (e.g., *go up*)? Why do Spanish speakers use a count noun in the plural to refer to popcorn (i.e., *palomitas*), where English speakers use a mass noun?
>
> *3. Consequence questions* ask about the effect of a particular grammar on other cognitive processes. For example, given that space is an integral component of the structuring of meaning in language, does being raised with two languages, each with their own way of representing spatial relationships, have any implications for the way that space is processed independently of language, as assessed by Shepard and Metzler's classic mental rotation task? One of the authors is currently engaged in addressing precisely this question.
>
> *4. Nondirectional relational* questions are typically used in correlational studies, where the investigator is not making conjectures about causality. Brain imaging techniques are often employed in an investigation of these questions, i.e., What parts of the brain lights up (are active) when people listen to a sentence describing active versus abstract sentences?

Each of these types of questions will yield qualitatively different answers. They each have a place in the investigation of any topic. Regardless of the type of question you choose, it must be translated into a research hypothesis.

6.1.5 *From research question to research hypothesis: scientific methodology*

Ayala (1994) describes the process of developing a research hypothesis as the most creative and imaginative element within research. The process begins, of course, with deciding on a research question, which is then interpreted as a research hypothesis. Interpreting your question as a scientific research hypothesis requires that you rethink it in terms of predicting what the relationship between two (or more) variables will be in a given context. Factors such as explanatory value and falsifiability also need to be integrated.

Earlier, scientific methodology was described as "an intellectual framework geared toward generating the most reliable findings possible as well as at facilitating their verification."

A more 'practical' definition of scientific methodology is:

> *Scientific methodology encompasses standardized methods of testing whether an idea, translated into a hypothesis, has explanatory value over a phenomenon in a setting that allows falsifiability.*

In other words, the next step is to frame your research question such that it makes predictions about the type of behaviors to expect whenever your variables (see below) interact in the ways you predicted such that your research question is addressed as directly as possible. In addition, your question must be framed such that the prediction can be shown to be wrong, meaning that it can be falsified.

A couple of examples should help clarify the process.

> *Example 1:* Casasanto and Boroditsky (2003, CS) are interested in how it is we understand concepts not directly available via direct sensory experience. This is their topic of interest and it is stated in the first few sentences of their paper. They also give the sources that started them thinking along those lines. Reading the work of researchers such as Gattis, Gibbs and Lakoff, etc. was instrumental in their choice of topic. They begin to narrow down their scope of interest by stating their interest in spatial metaphors for time. Notice that they give references to additional readings, i.e. Alverson, Boroditsky, etc. Their research question is whether the relationship between space and time, well-established in linguistic tasks, will also emerge in simple psychophysical tasks with nonlinguistic stimuli and responses. They re-interpret their research question as a research hypothesis centered on the asymmetry of the dependency between space and time appearing in language, postulating that it will also appear in psychophysical tasks.
>
> *Example 2:* For Bergen, Narayan and Feldman, the topic of interest is mirror neurons and their possible relationship with motion verbs. The sources that got them thinking about the issue include Gallese, Rizzolatti and Buccino. They are specifically interested in whether the comprehension of motions verbs involves some type of mental simulation, and whether this possible simulation might involve activation of the parts of the brain responsible for the behaviors described by the motion verbs. Further review of the literature led them to their research hypothesis. Namely, that information processed simultaneously by different modalities will occur at varying rates depending on its similarity.

Each of these sample papers begins with a broad statement of interest. The process resulting in a research hypothesis is well outlined by the literature they cite. Each step of the methodology is accompanied by supporting research. Though the whole process appears neatly packaged in the journal article, keep in mind that the authors spent a considerable amount of time reviewing different research, weeding out the useful from the irrelevant, all in an effort to develop as concise an image as possible of what the literature had to say about the topics they were interested in, as well as about what their research question should be and the form it should take as a research hypothesis. All the while keeping in mind the practicalities of actually finding out whether their idea could make a contribution or not.

6.2 Part two: The practical steps

6.2.1 *The methods section: Developing an experiment*

The methods section of a paper gives you the details of how a research hypothesis was tested. The authors will have arrived at this point only after having a very clear research hypothesis, as should you before beginning to consider the more practical steps. This is particularly important as you will next need to re-envision it within the context of the type of information an experiment can actually give you. Besides describing different research methods, this volume can also be thought of as a survey of the types of information produced by experiments. Common ones are response times, eye-movements, and differential activation in given areas of the brain. Deciding on a methodology involves deciding which type of information will best be able to address your research hypothesis. In order to do so, a clear understanding of research variables, experimental design, and differences between types of hypotheses will be necessary.

6.2.1.1 *Variables* Conducting a study, regardless of the type, is about attempting to understand how different components of a phenomenon, called variables, relate to each other. The most common types of relationships studied involve determining whether a variable is a part of a target phenomenon, or alternately, how it is relevant once it has been established that there is a connection.

6.2.1.2 *Variable classification* A variable can best be qualified as a set of events that can take on different values. Typical ones include sex, age, scores on an exam, number of milliseconds required to respond to a stimulus etc. They can generally be categorized in two different ways. Variables can be grouped according to their nature or according to how they are used in research.

Variables classified by their nature are behavioral, stimulus and subject (also sometimes called organismic). Behavioral variables comprise responses made by entities such as people or animals. How quickly a rat snatches a piece of chocolate is an example of a behavioral variable. Stimulus variables are those factors that precipitate a behavior. This might be a social context such as a party, or the reading of a sentence as a prompt. Subject variables are characteristics of objects of study that are used to classify them. They are generally not changeable. Sex and age are common examples.

Variables as classified by their use in research are Independent, Dependent, Extraneous and Constant. The independent variable is what is manipulated such that it will affect the dependent variable. Inversely, the dependent variable is what is affected by the independent variable. This relationship is unidirectional. For example, let us say that we were interested in finding out which social settings produced the most dancing. The independent variable would be the setting in which dancing took place, i.e. a birthday party, a wedding, an office meeting, a funeral. These are called levels of the independent variable, in this case 4 since there are 4 settings. The dependent variable would be the amount of dancing. The hypothesis is that the setting affects the dancing not the other way around, therefore, the independent variable is the setting and the dependent variable is the amount of dancing. It is important to keep in mind that there is no a priori reason why either setting or dancing should be the independent or the dependent variable. Their classification as a given type of variable is entirely dependent on the research hypothesis.

Extraneous variables are factors that might affect the independent variable. In the dancing examples above, alcohol might be considered an extraneous variable in that it might dispose subjects to dance more so than a particular social setting. To avoid alcohol from having any effect you control for it by making sure that no alcohol is consumed during the experiment.

What if you decided that you were in fact interested in how predetermined amounts of alcohol, i.e. 0–4 drinks, might affect the amount of dancing in each of the 4 settings? Then there would be 2 independent variables, the settings and different amounts of alcohol. In formal terms, what you would be interested in is the possible interaction between setting and alcohol consumption quantity.

By the same token, you can decide that you are interested in any physical reaction that might be indicative of response to the music at each of these settings (assuming that the music remained constant.) You could then operationalize (see below) behaviors such as foot tapping or swaying as levels of dancing, along with full body motion on the dance floor. This would give you a design (discussed further below) that has 2 independent variables, one with 4 levels, i.e. the 4 different social settings, and one with 5, i.e. the 5 different quantities of alcohol, beginning with no alcohol at all up to 4 drinks, and 1 dependent variable with 3 levels.

The last type of variable is a constant variable. These are factors that remain unchanged across the different conditions of an experiment. For example, a constant here would be to have the same number of subjects in each social situation. A further constant would be to have the same number of males and females. Constants are kept so as to prevent unwanted variation. It could be that if the number of subjects were different in each of the different dance conditions that the amount of dancing would vary as a result, i.e. different ratios of males and females induce more or less dancing. A constant can be thought of as a possible extraneous variable that has been controlled for.

6.2.1.3 *Internal structure of a variable* Regardless of how they are categorized, variables have internal structure. They can be discrete, i.e. the number of times a particular image is chosen in a forced-choice task, someone's gender, etc. or they can be continuous, i.e. the amount of force a subject applies to a lever, someone's age, etc. The key difference is that discrete variables are not decomposable, i.e. a family cannot have 1.5 children, while continuous ones are, i.e. today you ran for 22.37 minutes while tomorrow, if feeling more energetic you might run 44.06 minutes.

A variable's internal structure can also be described as qualitative or quantitative. The former refers to kinds, someone can be female, Chinese and left-handed. These are all kinds of things that she is, each of which has no intrinsic numeric value. Quantitative variables predictably refer to variables than can occur in different amounts. How much money is (or mostly likely, is not) in a graduate student's bank account, the amount of time she vacillates before calling up her parents for a loan, etc. are both examples.

6.2.1.4 *Operationalizing variables* When variables are used in an experiment, all target ones need to be operationalized. This refers to transforming any concept you are interested in into an observable behavior that can be measured. In the dancing example, we can operationalize 'dancing' as (attempted!) rhythmic movement performed by a person not including swaying while sitting or leaning against a wall, or foot-tapping while sitting or standing. A video recording might be made such that you can measure how much time people spent 'dancing' only according to the operationalized criteria. This way the amount of dancing will have an objective measure instead of being left to subjective interpretation, i.e. what may count as 'a lot of dancing' to one person may seem like not much at all to another. In addition, should it be the case that another experimenter doubts your results and wants to see for himself, he will know exactly what measures to use to test the reliability of your results. Proper operationalization is also basic to testing the theoretical concept you want to test. e.g., if you want to see how accessible a word is, you can operationalize this question in terms of how quickly people say a certain word (naming), or how quickly they determine it is a word in English (lexical decision). There is much debate on what measures best test a theoretical notion.

6.2.2 *Experimental design*

In the dancing experiment example above, the assumption was that music in given social settings would make people dance, and that alcohol consumption in moderate but increasing quantities would make them want to dance more. As described, the experiment is lacking two elements to qualify as a valid experiment. It requires a control group and random assignment.

Returning to our assumption, the implication is that people require a social situation of some type to want to dance, a desire that could be augmented by alcohol, i.e. no social situation equals no dancing. In order to control for the possibility that people might still dance when not in a social situation, i.e. when alone, 5 more groups will have to be created. In one group, one person will be alone listening to music during a period of time matching that of the social situations. The other 4 groups will also consist of people alone listening to music though they will receive the different quantities of alcohol. These conditions will serve to control for the possibility that situations involving music, though not social settings, might also induce dancing. Our design now involves a control condition involving no social situation, 4 social settings, and 5 alcohol consumption quantities. Since 'no social setting' can be considered a type of social setting, i.e. a non-social setting, we group it with social settings meaning that we have a 5×5 design with 3 levels of the dependent variable.

The last element is subject randomization. All this means is that all subjects are randomly assigned to each group. This is to prevent a particular group of people, students coming from a dance class, for example, to all be inadvertently assigned to the same group, thus biasing the results.

There are two general variations for the way this experiment can be conducted. If different subjects are assigned to one and only one group, then it is called a between subjects design because it means that when the analysis is done you will be making comparisons between groups. A stronger though more time consuming design is a within subjects design, although this has the possible drawback of informing all the subjects of the complete experimental design, and thus would require counterbalancing of conditions. Here, every subject is assigned to every group once, though in different randomized orders, each in turn used the same number of times. What makes this a strong design is that by putting every subject in every situation we can rule out the possibility that differences between groups were the result of unintended and uncontrolled differences between the subjects themselves. For example, the subjects in the birthday party setting who had 4 drinks might have just happened to like to dance a lot, regardless of the social setting or the amount of alcohol consumed. Since we will have data for every subject in every condition we can be more assured that the amount of dancing observed is the direct result of the social situation and the alcohol consumed, should alcohol prove to be an important factor, and not an extraneous variable.

6.2.3 *Research hypotheses and experimental hypotheses*

Whereas a research hypothesis makes predictions about a phenomenon, an experimental hypothesis makes predictions about the type of information that can be used to support a research hypothesis, i.e. these are hypotheses involving the experiments themselves. The basis for this distinction is practical. Most of the time it is impossible to observe or measure directly whether your variables will interact as predicted by the research hypothesis. For example, if the research hypothesis took the form of 'if A then B' using capitals, then the experimental hypothesis would be 'if a then b' with small letters. The catch is to make reasonably sure that the relationship between 'a' and 'b' maps to the relationship between 'A' and 'B.' The discussion section of a paper will typically address the effectiveness of the proposed mapping in detail. A large part of learning to be a critical reader involves learning to assess whether a researcher's claims about the research hypothesis are substantiated by the claims about the experimental hypothesis.

The methodology of your experiment will ideally present you with observable and measurable data that can help to support or negate the research hypothesis. This book is intended to introduce you to the different types of methods used to investigate language from different perspectives. Regardless of the type of question you ask, it is likely that the methods you find here will help you decide on the experimental protocol best suited to your question.

6.2.4 *The experimental and the null hypotheses*

Earlier we mentioned falsifiability and prediction. These two ideas are directly concerned with the development of hypotheses. The two simplest and also most common that you will encounter are 1) that the chosen task will produce the predicted effect or 2) that it will not. In this case they would be the experimental hypothesis and the null hypothesis, respectively. Prediction is involved with the experimental hypothesis since your hypothesis 'predicts' that given a certain setting, the given variables will interact in a certain way. The null hypothesis deals with plausibility. In essence, you cannot have a true experimental hypothesis that does not assume the possibility of being incorrect (Popper 1959). It is this possibility, which we call falsifiability, that is captured in the null hypothesis, i.e. that the effect you predict from the interaction of the given variables does not exist.

In actually conducting your experiment, the goal will not be to prove the experimental hypothesis correct but instead to show that the null hypothesis can be rejected. In other words, to establish whether your results merit rejecting the null hypothesis. Recall that earlier we described the scientific method as an attempt to come up with the best possible explanation for a phenomenon. This is achieved by showing that other explanations cannot account for the data. The null hypothesis encapsulates the other explanation as the proposal that the relationship predicted by the experimental hypothesis is not meaningful. If the relationship is shown to be meaningful using inferences based on statistical probability, then the new hypothesis explaining the relationship becomes the best available explanation for the phenomenon in question, thus rejecting the null hypothesis. If however, the proposed relationship is not shown to be meaningful, the null hypothesis is not rejected, thus maintaining the original explanation for the target phenomenon as the best available.

6.2.5 *Following are two examples of how all of these pieces fit together*

> *Example 3a:* Lozano & Tversky (CS, 2004) were interested in whether gesture benefits communication. Their research hypothesis is loosely that gesturing while explaining something benefits the person gesturing by facilitating reasoning, problem solving, and other cognitive processes. As an experimental method, they needed a controlled situation where people could be observed gesturing while explaining something. They decided to use a video-taping environment in which one person assembles something, then using their experience, reassembles the same object while explaining how to do it. A television cart was the object chosen, likely because it is relatively easy to assemble and disassemble. Their experimental hypothesis given this paradigm is that subjects who speak and gesture will reassemble the TV cart more accurately and efficiently than those who only gesture and a control group who will do neither. The null hypothesis is that gesture will have no effect on how well reassembly is done. The researcher's independent variable is the type of communication during reassembly, and it has 3 levels, speech + gesture, gesture only, or control. The dependent variable is the number of errors made during reassembly. Since the researchers want to know how much gesture benefits communication, counting the number of errors made as a direct result of a communicative instance is a very good measure of communicative effectiveness.

> *Example 4a:* Wiemer-Hastings, Barnard & Faelner (CS, 2004) were interested in the possible structural differences in abstract versus concrete categories. Their research hypothesis was that abstract and concrete item categories are structurally different because abstract items are organized more by situational relations, whereas concrete items are organized more by taxonomic relations. In their own words (p. 1453) "The main hypothesis was that abstract and concrete item categories differ in the amount of constraint that they place on membership." They needed an experimental method that would allow the membership flexibility of abstract categories, if real, to become readily apparent without manipulating the rigidity supposedly associated with concrete categories. They chose a listing paradigm in which subjects would list exemplars for each of 24 categories, half of which were concrete and half abstract. Their experimental hypothesis was then "that significantly more subjects would list the same items for concrete categories than for abstract ones" (ibid.). The null hypothesis was that there would be no difference in the listing for the two types

> of categories. Their independent variable was the use of either abstract or concrete items. Their dependent variables were the frequency of types produced and token/type ratio, i.e. a higher ratio is indicative of a category, which places strong constraints on cognitive processing.

There are two additional experiments in this paper. Take it as an exercise to break them down the way we have done with the others here.

6.2.6 *Reliability and validity*

Earlier we discussed the importance of learning to assess whether the research hypothesis and the experimental hypothesis mapped well to each other. There are two types of criteria instrumental in making this evaluation. They are reliability and validity.

Reliability refers to the consistency of the effect produced by your experimental method. If the test yields consistent results, it is considered reliable. Two common measures of the reliability of the experiment are test-retest and inter-observer. If an experiment is conducted on two or more separate occasions and the results are consistent, the experiment is considered to have met 'test-retest reliability'. In practice, most researchers will replicate, or conduct their experiments at least twice, to ensure that the results are trustworthy. A frequent factor interfering with test-retest reliability or even with reliability across subjects is inter-observer reliability. Reliability fails when different people performing different tasks for an experiment, i.e., collecting the data or coding it, do so differently across subjects or data sets. The goal of every investigator is to conduct every aspect of an experiment so that all tasks involved with it are done as consistently as possible so as to ensure that results are not made unreliable due to inconsistencies. This will be discussed further below in the section on data collection. Tests of reliability are assessed by the degree of correlation among measures (see Núñez for further discussion, this volume).

Validity primarily refers to how well your experiment measures what your research hypothesis says it measures. Validity can be assessed according to different criteria, and unfortunately, there are no statistical tests to measure it. Knowledge of the phenomenon in question is usually the only metric against which to assess the validity of the experiment. Common validity criteria are construct, ecological, internal, and external.

Construct validity is concerned with the operational definition of an abstract concept in an experiment. In our dancing example above, dancing was operationalized as full body rhythmic motion while on both feet, and not including foot tapping or swaying. Would you consider this a valid description of dancing? Can you think of a more exacting, less subjective metric?

A second kind of validity, ecological validity, involves how well an experimental design, and its results, fit with what people already know about their natural environment. It is sometimes difficult to see what a study on language as described in psychology journals really means for ordinary people in real life situations. Due to the necessity of a controlled environment, ecological validity is sometimes called into question. It is important that the investigator be cautious in understanding the trade-off between good experimental control, and doing work that is meaningful and informative outside of the laboratory.

The last two types are internal and external validity, and refer, unsurprisingly, to internal and external factors of an experiment. Internal validity concerns ensuring that the outcome of an experiment is not due to factors other than those intended by the experimenter. These can include elements such an uncontrolled extraneous variable, or inconsistency in data coding. External validity deals specifically with how well an experiment can generalize to the population at large. A common criticism of most university run research is that the subjects are mostly university students. This is a rather specialized population in terms of education, and sometimes socioeconomic status, among other possible factors. Deciding that findings from this group are applicable to other groups can sometimes be questionable.

Actually choosing a methodology to conduct an experiment means satisfying all of the criteria we have described. Once you have a research hypothesis, you must design an experimental protocol to address it. Choosing from among the methods described in this volume, or perhaps from a more closely related paper, you must identify all of your variables and your experimental hypotheses, as well as consider validity and reliability issues. Only then can you proceed to actual data collection.

6.2.6.1 *Following are the reliability and validity assessments for examples 3a and 3b*

Example 3b: Lozano & Tversky (CS, 2004)

Reliability assessment: In this experiment reliability could be assessed by replication. That is, if another group of subjects came into the lab and they were randomly assigned to the three groups, they would show similar results.

Validity assessment: The validity of this experiment rests on the assumption that reassembling a TV cart is a good measure of logical reasoning or problem solving.

Example 4b: Wiemer-Hastings, Barnard & Faelner (CS, 2004)

Reliability assessment: If another group of subjects were tested and showed similar results as the original group then it could be said that the measure is reliable. Do note that it will not matter if the abstract item categories themselves are variable. Reliability is dependent only on the consistent variability of the tokens for each category regardless of the group sampled.

Validity assessment: Validity in this experiment is the suitability of self-reported lists as a measure of mental organization. Is mental organization what is really being measured by these frequency data?

6.2.7 *Data collection*

There are three general aspects to data collection, obtaining subjects, actually collecting data from them, and ensuring their informed consent and safety.

6.2.7.1 *Collecting data* Although it is time-consuming and, for lack of a better term, grunt work, collecting data is a crucial step in the scientific process. In order for things to run smoothly, care must be taken to ensure that consistency reins throughout data collection. A few things to keep in mind follow.

It is usually best if the subjects do not know what your hypotheses are so as to prevent their affecting the results. A cover story is usually given to subjects to account for the procedure without giving away critical information.

It is also best if the person in charge of executing the experimental procedure does not know the hypotheses of the experiment. Experimenter bias can be a problem. We communicate with each other using means other than language. It is possible for an experimenter who wants a certain result to unconsciously signal the subject in some way. In order to prevent this, it is usually best to have naïve research assistants collect the data. If this is not possible, then the experimenter can try to assign subjects to groups randomly, so that the experimenter herself does not know which group they are in. This is called a double-blind technique. In addition, a strict protocol must be followed where instructions are read in the same way to all subjects, and careful attention is paid to treat all subjects consistently. Deviation can result in extraneous influences on the results.

6.2.7.2 *Obtaining subjects* University students tend to be the most common source of data. Many schools have a system in place to facilitate contact between researchers who want data, and students who will give it to them in exchange for course credit or a small monetary reward. There are other options for collecting data, however. And given the arguably valid criticism that comes from trying to extend findings from university-educated 18 to 25 year olds to the general population, it may be desirable to do so. Subjects can be obtained through advertising in local newspapers, by contacting community organizations, etc. A factor that will contribute to how easily this is accomplished will be transportable research equipment, i.e. people tend to be more willing to participate if they do not have to go out of there way very much to do so, thus if you can bring your equipment to them, you will have better luck. Alternately, the use of financial rewards may be required. Some studies requiring one observation per subject simply go on buses and subways and ask people one question.

6.2.7.3 *Subject informed consent and safety* The early history of human experimentation is filled with many exciting discoveries. Less well known is the history of mistreatment of experimental subjects. Much has been done in recent years to prevent any future abuse. All universities in the United States where research is conducted have human subjects experimentation review panels whose duty it is to oversee the conduct of all experimenters on their respective campuses. These panels follow guidelines set by the federal government.

As a researcher, your first priority will always be to ensure the safety and well-being of all subjects as well as to safeguard their rights and privacy, regardless of where the research is conducted. All practicing researchers learn early in their careers to only design experiments that adhere to these constraints. We cannot stress enough that there is no negotiation possible on this point.[2] Guidelines have been developed by the American Psychological Association, known as the Code of Ethics (http://www.apa.org/ethics/code2002.html,

2. Basic research must adhere strictly to these constraints. There are slightly different guidelines for cognitive research in medicine. The reasoning is simple. Basic research involves knowledge for its own

1992.) to help you plan your research. Though the guidelines go into extensive detail, the general ideas can be summarized as follows:

> *Well-being:* The experiment should not compromise the participant's well-being. This refers to experiencing unreasonable physical or emotional discomfort.
> *Informed consent:* Subjects must receive enough information about the experiment to be able to make an educated choice to participate. This, of course, does not mean that you should give away your hypotheses. It only means that subjects must receive information describing the procedure in enough detail such that they know what will happen. Subjects then sign a document describing the extent of their participation, and must receive a copy for their records.
> *Confidentiality:* Each participant's confidentiality must be safeguarded. This means that the only documents that can contain any identifying information must be their signed consent form and any possible subject list for the day. Identifying information must never appear in relation to any of the data collected. Additional information regarding the use of subjects for gesture and Sign Language data can be found in Wilcox & Morford, this volume.

The usual procedure for a new research project is that a proposal for research is submitted to the review panel. The proposal will typically contain a brief summary of the experiments to be conducted focusing on the treatment of subjects. It will be authorized for execution based on the acceptability of its procedures for the treatment of human subjects. Regardless of how innocuous you may believe your experiments to be, it is imperative to follow the requisite guidelines or you risk losing the privilege of conducting research.

6.2.7.4 *Additional points Conducting research off campus:* The requirements to collect data from persons who are not students at your home institution should be similar to those for using student subjects. The only difference will be that your identified subject source will not be students. If you are uncertain, it is best to contact your university's officials.

Conducting experiments with children: In the U.S., all research conducted on persons younger than 18 years of age, or in a university setting that may include 17 year old freshman, must be authorized by their legal guardian. This follows regardless of where the research is conducted. In other words, if you are based in the United States but plan to collect children's data in Uruguay, you will need to obtain written consent from the legal guardian of every child the same as you would if you were in the U.S. In terms of differences in age constraints, follow the strictest ones. For example, if your home country requires legal guardian consent for persons under 18 but your target country considers persons over 16 to be adults, follow the constraints of your home institution as these are the strictest.

Conducting experiments at different universities: The rule of thumb in conducting research at different universities is to investigate what the ethical requirements are at every location, well before you begin research, and to follow their guidelines to the letter.

---

sake. As such, no sacrifice from subjects should ever be expected. Medical research usually involves an acknowledged risk/benefit exchange.

These guidelines must be followed in addition to those of your home institution. If the university where you will be conducting research does not have a similar review panel, follow the most careful guidelines you can find. Well-respected research universities are a good source.

Conducting research in foreign countries: When conducting research in a foreign country, if you will be associated with a university, contact their human subjects review board the same as you would at your home institution. Otherwise, contact government officials for information on how to proceed. In countries where there are no set guidelines, be sure to contact the local government officials to ensure that you have all the necessary authorizations. In addition, the general protocol is that you must satisfy the guidelines of your home university as well as those of your research location.

6.2.8 *Data analysis*

Data analysis involves two types of statistical analyses. The first are descriptive and the second inferential (see Núñez this volume, for extended discussion.)

Descriptive statistics describe what the data look like using means, medians, modes, standard deviations and variances. The mean is the average of a set of data. For example, if you have 6 data points, {2, 4, 5, 6, 6, 9} the average would be 32/6 = 5.33. The median is the value that falls at the center of the data points if they were to be ordered from smallest to largest. If the data array has an odd number of entries it is the middle entry; if it is even numbered it is the mean of the two middlemost entries, (5 + 6)/2 = 5.5. The mode is the largest cluster of similar values. In this case it would be 6, because there are two of them and only one of each other value.

Of these three, the most commonly used is the mean. The standard deviation and the variance depend on this number. They both measure the spread of your data. The standard deviation is (loosely!) the average value by which each data point deviates from the mean. It is calculated by subtracting the mean from each data point, producing that data point's deviation, then squaring it, then doing the same for all data points, then adding the sums of all of the squares of all of the deviations for all of the data points, dividing by the number of data points minus one, i.e. n-1, then taking the square root of that total. The chart below should make this much clearer. The formula is $s = \sqrt{\frac{\sum(x-m)^2}{n-1}}$

Data point (x)	Mean (m)	Deviation ($x - m$)	Squared deviation $(x - m)^2$
2	5.33	–3.33	11.0889
4	5.33	–1.33	1.7689
5	5.33	–0.33	0.1089
6	5.33	0.67	0.4489
6	5.33	0.67	0.4489
9	5.33	3.67	13.4689
		Sum of the squared deviations = $\sum(x - m)^2$ =	27.3334
		Divided by $n - 1(6 - 1)$ =	5.46668
		Squared root =	2.34

Here it is 2.34. In general, the smaller the standard deviation, the more consistent your data set. Usually the standard deviation is transformed into a variance by squaring it. Given our standard deviation, the variance here is 5.48. The variance is preferred to the standard deviation as a measure of the variability in a data set because it is considered nicer from a mathematical perspective, however both are used in inferential statistics, discussed below. For our purposes, both measures are useful.

Once you have calculated the descriptive statistics for each of your groups, the next step is to begin comparing them. The key question while exploring your preliminary results is whether there are noticeable differences between the different sets. This is typically done by graphing them in different ways, i.e. using bar-charts, pie charts, box and whiskers plots etc. What you are looking for is visual evidence of differences between your groups, as well as patterns connecting the differences. Returning to the dancing example above, graphing the data might reveal that the experimental condition where the most dancing occurred was the birthday party where subjects each had 2 drinks. At least, in looking at a bar chart of the data, this bar seems to be that largest. Testing whether this finding is significant requires inferential statistics.

Inferential statistics are used to show whether the differences you see between the different groups are meaningful enough to reject the null hypothesis. Though there are different ways that this can be done, they all depend largely on probability theory, or the likelihood that your findings are the result of chance. The details of these procedures are complex enough that they have been given their own chapter. Núñez in the following chapter, will provide you with a detailed description of how to proceed.

6.2.9 *Interpretation*

The ideal situation is that your results turn out exactly the way you predicted. In this case, all that is required is a formalization of your findings in the form of a research article. Fortunately, human cognition is far too interesting and complex to make things that easy. In fact, the results of an experiment rarely turn out exactly as the design predicted. For example, what if a group of researchers wanted to know if there was a relationship between hair color and height. The researchers recruited 60 men and women, and assigned them to either the "tall" or "short" group. They find that there are significantly more brown-haired people in the short group than the tall group. Thus, it could be concluded that there is a relationship between brown hair and height. However, there is also a significant interaction of gender by hair color by height. What should be reported?

First, the researchers should report all significant interactions. However, if you have a 5-way design, then there are 20 interactions making this rule less than practical. Often people do not report the higher level interactions simply because they are uninterpretable. (However, there are cases where reviewers or readers will find significant effects interesting or even interpretable, despite the author's initial classification as irrelevant.) Though it is a bit of a gray area, the authors should report unusual, or unexpected, significant results within reason. It may then be necessary for them to go back and look for possible causes of the effect in their data, a flaw in how the data was recorded or a factor they had not originally considered. A careful check should be done of how people were assigned to groups to ensure that it was done without bias. In other cases, where a computer or other recording apparatus was used, it should be checked for recording error as well. Once these checks are in place, the researchers must treat the effect as real, and offer an interpretation with their report. In this case, it may be that the tall men are mostly blond, whereas the number of blond women do not differ in either the tall or short group. It can be said that "the men carry the effect" in the resulting relationship between brown hair and height. And it is the reason why there is a significant interaction.

Such an explanation might seem obvious in this example, however, as we discussed earlier it is necessary at various stages of the scientific process to revisit the literature. At this point any explanation is post hoc, which means that the researchers did not predict this effect, but given its emergence, can make a logical argument for it. Post hoc explanations are not very convincing to scientists, who would be more convinced by a second experiment. There is a silver lining to unforeseen effects in that they will sometimes lead to further experimentation which can either be set up to have better controls, or modulate the effect of the newly considered variable.

7. Then there is the rest of the planet...

Cognitive linguistics (CL), perhaps more so than any other discipline encompassed by cognitive science, has an implicit interest in cross-linguistic research. For CL, embodiment is a basic premise. Language is not a complete abstraction. Instead it is grounded in the experience of our bodies in all of the environments they inhabit, from the physical to the social, to everything in between. In other words, CL is based on the idea that language is an extension of our environments. If our environments differ, then likely so will our languages. To better understand how this process works, it is necessary to study and compare many languages in conjunction with the cognitive systems they are a part of.

The initial sections of this chapter present an experimental environment that more and more researchers recognize as an idealization. The possible effect of factors such as cultural or linguistic differences and abilities, have not usually been considered. Nonetheless, evidence keeps emerging supporting cognitive sensitivity to these influences.

A telling example from research on the cultural basis of cognition comes from Cohen and Gunz (2002). When the authors investigated differences in memory recollection between westerners and easterners, they found that when easterners recalled a memory of an event of which they had been the focus, they spoke about themselves in the third person, but when they had not been the focus, they spoke about themselves in the first person. This is exactly the opposite of what would be expected in the west where the opposite is the norm. According to the authors, these differences in perspective indicate deeper differences in the way that information is processed and in the way that memories are eventually encoded.

Other work explores the ways that specific languages mold cognition. Levinson's (1996) work comparing Tzeltal and Dutch showed that the system used to describe spatial relationships in a language had consequences for the way that space was understood in non-linguistic processes. Languages tend to use either a relative location (left/right, front/back) or absolute location system, akin in meaning to the English cardinal direction system, i.e. north, south. Dutch uses a relational location system and Tzeltal an absolute one. Levinson developed a task that involved comparing arrows in different orientations set on two different tables. Subjects were asked to point to the arrow in the second table that matched the direction of the arrow in the first. Whereas Dutch speakers tended to choose the arrow that matched the original direction relative to themselves, Tzeltal speakers chose the one that matched the absolute direction of the first, a result which mapped to the spatial system used in the subjects' languages.

Much as we would like to follow with another item by item set of instructions for how to deal with cultural differences in experimentation, this is impractical given not only space considerations, but the sheer impossibility of preparing for every possible scenario. The best we can do is to help make you aware of the issues. They center on three foci, types of human groupings, linguistic variation and endemic bias in the interpretation of cognitive phenomena.

7.1 Types of human groupings

People tend to define themselves according to the groups they belong to as much as by those they do not. Common groups include 'culture,' 'race,' 'ethnicity,' 'nationality,' and 'socio-economic status.' If addressed directly, most people easily indicate those they belong to. The problem is that there are no a priori agreed upon constraints for who belongs to which group. Some people might choose to classify themselves as belonging to a certain group using completely different criteria than other people. In terms of race, for example, in some countries, having any African ancestry qualifies someone as belonging to the African race. In others, having any European ancestry designates them as white. The classification is thus entirely subjective and tied to a given set of societal norms. All human groupings are subject to such constraints. As a researcher, you will find yourself invariably trying to determine how to best select your subjects so as to minimize effects from possible external variables. This will be important whether you are interested in contrasting different groups or focusing on a phenomenon in as generic an environment as possible. There is no easy answer for how to control for grouping factors. The best you can do is to be aware of their nature so as to be better able to control for them, and to document them as completely as possible in your research report. The following descriptions of some of the main groupings are intended to introduce you to the issues involved in making useful classifications.

7.1.1 *Culture*

What is culture? A textbook definition is "the customs, behaviors, attitudes, and values, and the objects and implements that can be used to identify and characterize a population (Beins 2003: 309). Though this description might seem uncontroversial enough, following is a scenario that should help you better conceptualize the issues. Say you think of yourself as an American, for example, visiting a country different from your own, say Canada. Your 'culture' as an American is probably very similar to that of Canada. Or is it? Though there are countless similarities between the two cultures, there are differences. Canada has a significant social safety net and much lower levels of violent crime, for example. Are these differences significant enough to, in turn, merit cognitive differences?

7.1.2 *Race*

From a scientific perspective, there is no such thing as different human races. We are one race. However, despite the ivory tower we may sometimes choose to find refuge in, we live in a deeply politicized world. Many governments assume racial distinctions based on differences in physical morphology when collecting census data. A difficult task complicated by the fact that many people are "mixed." Given the absolute lack of any biological evidence of anything but cosmetic differences between different people, this is not likely to be an attribute that will need to be dealt with directly in research. However, physical morphology is often correlated with other types of groupings. Perhaps the most common is ethnicity, described below.

7.1.3 *Ethnicity*

Ethnicity, as used in the United States, refers to cultural heritage (Shiraev & Levy 2001). Typically this includes religion, language, ancestral origin, traditions, diet, etc. Belonging to the same ethnicity does not entail having the same national origin. There are ethnic Hungarian communities in Slovakia, Austria and obviously in Hungary, among other countries, for example. By the same token, people that might be grouped as belonging to the same ethnicity by a given country's sociopolitical structure do not necessarily share more than language and possibly religion. Mexican and Puerto Rican people in the United States are a case in point. Though both original cultures were colonized by the Spanish, resulting in the adoption of the same language and religion, most similarities end there. Puerto Rican culture resulted from the mixing of surviving Taino, enslaved and imported Africans and colonizing Spanish. Mexican culture resulted mostly from the mixing of established though subjugated Native American societies and colonizing Spanish. The subsequent extant cultures are hybrids of the originals, and therefore, though superficially similar, are actually quite different.

7.1.4 *Nationality*

Nationality can be just as confusing as ethnicity. Though the dictionary definition describes a group of people who share history, language, geographical origin and exist under the protection of a recognized and formalized political entity, i.e. a country, people belonging to the same nationality can be quite diverse. Someone can have emigrated from Iran, and have been naturalized as a United States citizen. Another person might belong to a family that has lived in the United States for 8 generations, having originated from England. Yet a third might be a direct descendent of the people the Pilgrims encountered when they first arrived. Nationally, all of these people are considered Americans.

7.1.5 *Socio-economic status*

Socio-economic status usually involves a grouping based on family income, parental educational level, parental occupation, and social status in the community. Here again there are a number of inconsistencies that can arise due to intergenerational differences and emigration. A member of a family that is well off, could for many reasons, work in a fast food restaurant and barely make enough money to live. A person whose parents were subsistence farmers could receive a scholarship and become a well-educated high-paid engineer as an adult despite having been raised in poverty. An immigrant who was a teacher in his home country can find himself cleaning bathrooms upon arriving in his adopted country.

How do you decide which of these groupings to use as subject characteristics and how to use them effectively? Once again, there is no neat answer. It really does depend on the specifics of the phenomenon you are investigating. Further, perhaps you have noticed that all of the distinctions of types of groups are based on the constraints of the English-speaking world. There is no guarantee that the people you might be interested in studying would necessarily use them. At the risk of sounding like a dangerous stunt television show, a strong recommendation is to not attempt to do a study comparing these factors either as a first study or alone if you have no experience with studies in general. Science is a strongly collaborative endeavor. Enlist the help of someone with at least some experience. That said, the best you can do, once again, is to document as completely as possible the choices made, as well as to be as consistent as possible with the internal structure of your sample.

7.2 Linguistic ability

Bilingualism and multilingualism in general present another set of variables. If your direct goal is to investigate cognitive differences between groups with differing linguistic abilities, you should obviously develop criteria for inclusion and exclusion of subjects into your various study groups. Though the required guidance is beyond the scope of this volume, the "Handbook of Bilingualism" edited by Kroll and De Groot (2005) will prove a useful guide.

Chances are that your study will not focus on multilingualism and you may wonder how relevant the phenomenon might be to your research. Consider the following recent studies.

Marian and Spivey (2003a, b) conducted a series of eye-tracking studies on phonological interference. They tested Russian-English bilinguals to see if the phonological attributes of one language would interfere with those of the other. They found that when subjects were presented with an array of objects whose names overlapped phonetically with each other, albeit in different languages, i.e. "marker" in English versus 'marka' in Russian meaning stamp, and asked to point to an object, for example, the marker, that they also made eye-movements toward the stamp. According to the authors this indicated phonological interference from the second language.

In a study on short-term memory, Thorn and Gathercole (1999) showed that when bilingual, monolingual, and ESL students were tested on a word-recall task, the number of recalled words was a function of how well the subject knew the language the words belonged to, indicating that short-term memory is not a language-independent process.

Emerging evidence shows what appear to be differences in non-language related cognitive abilities in bilinguals. Bialystok (1992, 1994, 1999, 2005) has shown that bilinguals exhibit greater control of attention in the execution of cognitive tasks. In other words, that they are better at identifying and concentrating on the relevant elements of a task while effectively ignoring peripheral details. In terms of spatial abilities, Bialystok and Majumder (1998) showed that bilinguals were favored over monolinguals in spatial control tasks. Further, McLeay (2004) showed that bilinguals were both faster and more accurate in a mental rotation task involving knots.

These are all examples of ways that multilingualism produces cognitive differences. As a responsible researcher, the course of action to address the possibility that your results might be influenced by variations in linguistic ability is the same it would be with any other variable. Control for it. This means ensuring that your subjects all belong to the same linguistic group or else documenting any within-subjects variation. That way, if you do get results that are not quite what you expected, you can use statistical analyses to rule out those effects. This is precisely the same technique used for any other subject variable not the direct object of investigation. Sex, for example, is commonly documented for each subject. Researchers routinely check to see if there are any significant differences in the results attributable to sex. Linguistic variation should be treated in much the same way.

8. Endemic bias in the interpretation of cognitive phenomena

Earlier, in describing the scientific method, we talked about the inherently subjective nature of any human endeavor. Often, subjectivity takes the form of biased thinking of which people are largely unaware. Shiraev and Levy (2001: 55–70) provide an excellent description of common biases in conducting research between different groups. Following is a brief summary of the main points along with the antidotes for each bias proposed by the authors.

8.1 The evaluative bias of language

Words are meant to describe the entities in our environments. What we do not realize is that they also prescribe what they are. If I call a flat surface suspended on four legs a table, then it 'is' a table, if I call it a desk, it becomes a desk. In terms of the way that people are classified, consider these pairing, all of which are meant to refer to the same characteristic. *Old/mature; obsessed/committed; lunatic/visionary; dead/ontologically impaired.* These terms each carry value judgments which are impossible to escape, i.e. one term is considered more positive and the other more negative. Try to come up with a non-value laden description of *narcissistic/high self esteem*! Scientific language is not immune to biasing, whether it be in describing a target phenomenon or writing a consent form.

Antidotes:

1. Remember that descriptions, especially concerning personality characteristics can never be entirely objective.
2. Become aware of your own personal values and biases, and how they influence the language you use.
3. Avoid presenting your value judgments as objective reflections of truth.
4. Recognize how other people's use of language reveals their own values and biases.

8.2 Differentiating dichotomous variables from continuous ones

There are phenomena in our environment that occur as mutually exclusive or contradictory pairings. A woman can be pregnant or not, but she cannot be a little pregnant. Many other phenomena are treated as if they were also dichotomous, even though they are actually continuous. Examples are cooperative/competitive; introverted/extroverted; good/bad. This can be a problem, for example, while designing an experiment in that a variable can be inadvertently treated as if it were dichotomous, when in fact it is continuous.

Antidotes:
1. Learn to differentiate between variables that are dichotomous and those that are continuous.
2. Remember that most person-related phenomena – such as traits, attitudes, and beliefs – lie along a continuum.
3. When making cross-cultural comparisons, try to avoid artificial or false dichotomies.

8.3 The Barnum effect

The Barnum effect refers to statements that are generally true about most people. It references Barnum in that it has ‘a little something for everyone.’ Barnum statements treat the general as if limited to the specific. One could say, ‘women are sensitive to sexism’ which seems true though upon deeper thought is also likely to be true of men.

Antidotes:
1. Learn to differentiate Barnum statements from person and group specific descriptions and interpretations.
2. Be aware of the limited utility inherent in Barnum statements. Specifically, remember that although Barnum statements have validity about people in general, they fail to reveal anything distinctive about any given individual socio-cultural group.
3. Whenever feasible and appropriate, make it a point to reduce the Barnum effect by qualifying personality descriptions and interpretations in terms of their magnitude or degree.

8.4 The assimilation bias

We categorize everything in our environments as a matter of course. Typically, not much thought is involved when we use schemas or cognitive structures to organize our beliefs. Though they are useful to a point, i.e. you are a woman alone walking on a street when you see three men walking in your direction and so you immediately assume they might be dangerous and cross the street, they can be problematic when they function as unquestioned stereotypes, the three men were actually your colleagues from work whom you have now insulted by so obviously avoiding them. In short, what schemas do is help you make assumptions about a person or situation based on limited information. They are a problem when we do not see beyond them.

Antidotes:
1. In situations in which you are likely to utilize the representative heuristic, make a conscious effort to consider the possibility that the schema or prototype in question might be inaccurate, biased, or incomplete.
2. Take into account relevant statistical information, such as base rates, sample sizes, and chance probability.
3. Beware of the natural tendency to overestimate the degree of similarity between phenomena and categories.
4. Recognize that your personal attitudes about people and group prototypes can bias your comparison and subsequent judgments.

8.5 Fundamental attribution error

This error refers to the tendency to assume that 'cause' of someone's behavior is their character and not the situation. In essence, behavior is attributed to internal influences versus external ones. For example, you go to a crowded restaurant and are served by a waiter that responds brusquely and is late with your order. You may assume that the person is a bad waiter, and dismiss him at that. What you may not know is that the kitchen is in disarray, and that the cook just yelled at the waiter for no reason other than that he was the first person to walk through the door. If the waiter is actually usually quite courteous and timely, in categorizing him as a bad waiter you have just committed fundamental attribution error by not considering that the situation might also have a role to play in his behavior.

Antidotes:
1. Do not underestimate the power of external, situational determinants of behavior.
2. Remember that at any given time, how people behave depends both on what they bring to the situation as well as on the situation.
3. Keep in mind that this attributional error can become reversed, depending on the perceiver's point of view. Specifically, although people are prone to underestimate the impact of others' situations, they tend to overestimate the impact on their own situations.
4. Be sure to take into account both cognitive and motivational biases that are responsible for producing these attributional errors.

8.6 Correlation does not prove causation

This is actually one of the most common errors made in considering the relationship between two phenomena. We assume that because two events happen at approximately the same time, that somehow, one 'caused' the other. For example, there is some evidence that watching violent television programs is correlated with violent behavior. Some people have interpreted this to mean that watching violent television programs makes one violent. The possibility that violent people might like watching violent programs is not considered, thus reinterpreting the correlation not as causative but as symptomatic of otherwise violent behavior.

Antidotes:
1. Remember that correlation or coappearance is not, in itself, proof of causation.
2. Keep in mind that correlations enable us to make predictions from one event to another; they do not , however, provide explanations as to why the events are related.
3. When a correlation is observed, consider all possible pathways and directions of causation. For example, if Event A and Event B are correlated, does A cause B? Does B cause A? Do A and B cause each other? Does C cause A and B?

9. Conclusion

The goal of this chapter was to get you on your feet, so to speak, about experimentation. It would be naïve to imagine that all you need to get started can be found here. As social creatures we tend to learn best if we have a model of how to proceed. If at all possible, in lieu of going it alone, try to find someone who would be willing to serve as your mentor during this process. At the very least, seek out someone who will allow you to observe as they develop a project through to the point of writing up the research paper. The experience will provide you with a valuable concrete model of how to conduct your own work.

There are also excellent methods and statistics books available to help you further develop your knowledge of the material introduced here. We strongly recommend those referenced here as well as Geoffrey Keppel's "Design and Analysis: A Researcher's Handbook" in whatever edition you can find it in, as well as Julian Meltzoff's (1998) "Critical Thinking about Research: Psychology and Related Fields."

References

American Psychological Association (1992). Code of ethics (http://www.apa.org/ethics/code2002.html)

Aronson, E., Wilson, T. W., & Marilyn B. Brewer (1998). "Experimentation in social psychology." In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology, Volume 1, 4th/ed*. Oxford: Oxford University Press.

Ayala, F. J. (1994). On the scientific method, its practice and pitfalls. *History and Philosophy of the Life Sciences 16*, 205–240.

Beins, B. (2003). *Research methods: Tools for life.* Boston: Allyn & Bacon.

Bergen, B., Narayan, S., & Feldman, J. (2003) Embodied verbal semantics: Evidence from an image-verb matching task. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Bialystok, E. (1992). Attentional control in children's metalinguistic performance and measures of field independence. *Developmental Psychology, 28*(4), 654–664.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16*(2), 157–168.

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*(3), 636–644.

Bialystok, E. (2005). The impact of bilingualism on language and literacy development. In Kroll & De Groot (Eds.), *Handbook of Bilingualism.* New York: Oxford University Press.

Bialystok, E. & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*(1), 69–85.

Bloom, P., Peterson, M. A., Nadel, L., & Garrett, M. (1996). *Language and Space.* Cambridge: MIT Press.

Carnap, R. (1966). *The Philosophy of Science.* New York: Basic Books.

Casasanto, D. & Boroditsky, L. (2003). Do we think about time in terms of space? *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Chalmers, A. (1976). *What is this Thing Called Science?* Queensland: Queensland University Press and Open University Press.

Cohen, D. & Gunz, A. (2002). As seen by the other… Perspectives on the self in the memories and emotional perceptions of Easterners and Westerners. *Psychological Science, 13,* 55–59.

Ember, C. R. & Ember, M. (2001). *Cross-cultural research methods.* New York: Altamira Press.

Jordan & Zanna (1999). How to read a journal article in social psychology. In R. F. Baumeister (Ed.), *The Self in Social Psychology* (pp. 461–470). Philadelphia: Psychology Press.

Keppel, G. (1994). *Design and Analysis: A Researcher's Handbook.* Englewood Cliffs, NJ: Prentice Hall.

Kroll & De Groot (2005). *Handbook of Bilingualism.* New York: Oxford University Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the Growth of Knowledge.* Cambridge, MA: Cambridge University Press.

Levinson, S. (1996). Language and Space. *Annual Review of Anthropology*, *25,* 353–382.

Lozano, S. C. & Tversky, B. (2004). Communicative Gestures Benefit Communicators. *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*.

Marian, V. & Spivey, M. (2003a). Comparing bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193.

Marian, V. & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*, 1–19.

McLeay, H. (2004) The relationship between bilingualism and spatial tasks. *Bilingual Education and Bilingualism, 6*(6), 423–438.

Julian Meltzoff's (1998) *Critical Thinking about Research: Psychology and Related Fields.* Washington: American Psychological Association.

Newcombe, N. S. & Huttenlocher, J. (2000) *Making Space: The Development of Spatial Representation and Reasoning.* Cambridge, MA: MIT Press.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Rosnow & Rosenthal (1993). *Beginning behavioral research: A conceptual primer.* New York: Macmillan Publishing Company.

Shiraev, E. & Levy, D. (2001). Introduction to Cross-Cultural Psychology. Needham Heights: Allyn and Bacon.

Solso, R. L. & Johnson, H. H. (1984) An Introduction to Experimental Design. In *Psychology: A Case Approach* (3rd ed.). New York: Harper & Row.

Sternberg, R. (2003). *Cognitive psychology* (3rd ed.). Belmond, CA: Wadsworth.

Tattersall, I. (2002). The Monkey in the mirror: Essays on the science of what makes us human. New York: Harcourt.

Thorn, A. S. C. & Gathercole, S. E. (1999). Language-specific Knowledge and Short-term Memory in Bilingual and Non-bilingual Children. *The Quarterly Journal of Experimental Psychology A, 52*(2), 303–324.

Wiemar-Hastings, K., Barnard, K. K., & Faelnar, J. (2004). Structural differences in abstract and concrete item categories. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Inferential statistics in the context of empirical cognitive linguistics

Rafael Núñez

1. Introduction

Imagine you read in a book about English grammar the following statement:

> The regular plural marker in English is 's', which is placed at the end of the noun as in *dog/dogs*. There are, however, some exceptions like *mouse/mice*.

Then, you read in a book about child development a statement such as this one:

> English speaking children have difficulties learning irregular forms of the plural.

The two statements make reference to a common subject matter, namely, English plurals, but they reflect deep differences in focus and method between the fields they belong to. The former statement is one about regularities and rules of a *language* (English), while the latter is a statement about patterns and regularities in the behavior or performance of *people* who speak (or learn) that language. The latter is about phenomena that take place at the level of *individuals*, while the former is about phenomena that lie *beyond individuals* themselves. The former is the purview of linguistics, the latter, of psychology.

As an academic discipline one of the main goals of Linguistics is to study patterns and regularities of *languages* around the world, such as phonological structures and grammatical constructions like markers for tense, gender, or number, definite and indefinite articles, relative clauses, case marking of grammatical roles, and so on. Psychology, on the other hand, studies people's behavior, motivation, perception, performance, cognition, attention, and so on. Over the last century, the academic fields of linguistics and psychology have developed following rather different paths. Linguistics, by seeing its subject matter as existing at the level of structures of languages – beyond individuals themselves – gathered knowledge mainly through the careful and detailed analysis of phonological and grammatical patterns, by means of what philosophers of science and epistemologists call the *method of reasoning* (Helmstadter 1970; Sabourin 1988). Psychology, on the other hand, by marking its definitive separation from Philosophy during the second half of the 19th Century, defined its subject matter at the level of the individual and embraced a method of knowledge-gathering which had been proved useful in the hard sciences: the *experimental method*. Ever since, peoples' behavior, performance, and mental activity began to be studied with rigorously controlled experimental methods of observation and data gathering. An essential tool of the modern experimental method is *Statistics*, especially inferential statistics, which is a carefully conceived mathematically-based conceptual apparatus for the systematic and rigorous analysis of the numerical data researchers get from their measurements.[1] The main goal of this chapter is thus to characterize some key concepts in statistical reasoning in the context of empirical cognitive linguistics, and to analyze how inferential statistics fits as a tool for knowledge gathering in cognitive linguistics. The text is not meant to be a comprehensive introduction to statistics proper, for which the reader may consult any college textbook available, such as Witte & Witte 2007; Hinkle, Wiersma, & Jurs 2003; Heiman 2003, but is instead intended to introduce the topic in a way that is meaningful for cognitive linguists.

2. What counts as "empirical" in cognitive linguistics?

One important difference between linguistics and psychology has to do with the question of what counts as *empirical evidence* for falsifying or supporting a given theoretical claim. What is a well-defined valid piece of information that can be safely incorporated into the body of knowledge in the field? What counts as an acceptable method for obtaining such a piece of information? When can that piece of evidence serve as a robust counterargument for an existing argument? These are questions for which linguistics and psychology traditionally have had different answers. For example, if some theory in linguistics states that

> the plural marker in language *X* is provided by a marker *x*,

then if through your detailed observations of the language you find a counterexample where the plural marker is not *x* but *y* (i.e., well-accepted linguistic expressions that do not fit the proposed pattern) you can, with that very piece of information, falsify the proposed theoretical statement. In such a case you would be making a contribution to the theory after which the body of knowledge would be modified to become something like

> the *regular* plural marker in language *X* is provided by a marker *x*, but there are some *irregular* cases such as *y*.

The same would happen if some linguistic theory describes some grammatical pattern to be generated by a rule *A*, but then you observe that if you apply such a rule to some sentences you actually generate *ungrammatical* ones (which linguists usually write down prefacing it with an asterisk). Again, by showing these ungrammatical sentences you would be providing linguistic *evidence* against a proposed theoretical statement. You would be falsifying that part of the theory and you would be de facto engaging in a logical counterargumentation, which ultimately would lead to the development of a richer and more robust body of knowledge in that field.

1. The history of Statistics as a field, however, reveals that its development not always came straightforwardly from advances in mathematics (see Stigler (1986) for details).

The situation in psychology is quite different. Because the subject matter of psychology is not defined at the level of the abstract structure of a language as linguistics does, but defined at the level of peoples' behavior (often used to infer people's cognitive mechanisms), performance or production, the elements that will allow you to falsify or support a theoretical statement will be composed of empirical data observed at the level of peoples' behavior or performance. For instance in order to falsify or support a theoretical statement such as

English speaking children have difficulties learning irregular forms of the plural

you would have to generate a series of studies that would provide empirical data (i.e., via observation of real children), supporting or challenging the original statement. Moreover, you would have to make several methodological decisions about how to test the statement, as well as how exactly to carry out your experiments. For example, because you will not be able to test the entire *population* of *all* English speaking children in the world (much less those not yet born!), you will have to choose an appropriate *sample* of people who would represent the entire population of English speaking children (i.e., you would be working only with a subset from that population). And you will have to *operationalize* terms such as "difficulties" and "learning" so it is unequivocally clear how you are going to *measure* them in the context of your experiment. And because it is highly unlikely that all individuals in your sample are going to behave in exactly the same manner, you will need to decide how to deal with that *variability*, and how to characterize in some general but precise sense what is going to be considered a "representative" behavior for your sample (i.e., how to establish a proper measure of *central tendency*), how to describe the degree of similarity of children's behavior, how to objectively estimate the error involved in any measurement and sampling procedure, and so on. Within the realm of the experimental method all these decisions and procedures are done with systematic and rigorous rules and techniques provided by statistical tools.

The moral here is that linguistics and psychology, for historical and methodological reasons, have developed different ways to deal with the question of how you gain knowledge and build theories, how you falsify (or support) hypotheses, how you decide what counts as evidence, and so on. Where the relatively new field of cognitive linguistics is concerned, which gathers linguists and cognitive psychologists (usually trained within the framework of the experimental method), it is very important to:

(a) keep in mind the nature and the implications of these methodological differences, and
(b) understand the complementarity of the two methods of knowledge-gathering as practiced in linguistics and experimental psychology.

Cognitive linguistics, emerging out of linguistics proper, and initially in reaction to mainstream Chomskian-oriented approaches, has made important contributions to the study of human language and its cognitive underpinnings. This new field, however, relying mainly on linguistic methods of evidence gathering has made claims not only about languages, but also about the psychological reality of peoples' cognition. For example, an important subfield of cognitive linguistics, conceptual metaphor theory, has, for the last twenty-five years or so, described and analyzed in detail thousands of metaphorical expressions such as *the teacher was quite cold with us today* and *send her my warm helloes* (Lakoff & Johnson 1980; Lakoff 1993). The inferential structure of these collections of linguistic expressions has been modeled by theoretical constructs called *conceptual metaphors,* which map the elements of a source domain (corresponding in the above examples to the thermic bodily experience of Warmth) into a more abstract one in the target domain (in this case, Affection). The inferential structure of the metaphorical expressions mentioned above (and many more) is thus modeled by a single conceptual metaphor: Affection Is Warmth.[2]

But beyond the linguistic description, classification, and analysis of *linguistic expressions* using methods in linguistics proper, conceptual metaphor theory has made important claims *about* human cognition, abstraction, and mental phenomena. For instance, it has claimed that conceptual metaphors are ordinary inference-preserving *cognitive mechanisms*. These claims, however, are no longer at the level of linguistic data or the structure of languages, but at the level of individuals' cognition, behavior, and performance. Because of this, many psychologists believe that giving a list of (linguistic) metaphorical expressions as examples does not provide evidence (in cognitive and psychological terms) that people actually *think* metaphorically. In other words, what may count as linguistic evidence of metaphoricity in a collection of sentences for linguists may not count as evidence for psychologists that metaphor is an actual cognitive mechanism in peoples' minds. It is at this point where some psychologists react and question the lack of empirical "evidence" to support the psychological reality of conceptual metaphor (Murphy 1997; Gibbs 1996; and see Gibbs in this volume). Moreover, how do we know that some of the metaphors we observe in linguistic expressions are not mere "dead metaphors," expressions that were metaphorical in the past but which have become "lexicalized" in current language such that they no longer have any metaphorical meaning for today's users? How do we know that these metaphors are indeed the actual result of real-time cognitive activity? And how can we find the answers to such questions? These are reasonable and genuine questions, which from the point of view of experimental psychology, need to be addressed empirically.

Cognitive linguistics, which is part of cognitive science – the multidisciplinary scientific study of the mind – is located at the crossroads of linguistics and cognitive psychology and as such, has inherited a bit of both traditions. It is therefore crucial for this field that the search for evidence and knowledge be done in a complementary and fruitful way.

We are now in a position to turn to more technical concepts.

3. Descriptive vs. inferential statistics

Statistics has two main areas: *Descriptive* and *inferential.* Descriptive statistics, which historically came first, deals with organizing and summarizing information about a collection of actual empirical observations. This is usually done via the use of tables and graphs (like

2. In conceptual metaphor theory usually the name of the mapping has the form *X* Is *Y*, where *X* is the name of the target domain and *Y* the name of the source domain.

the ones you see in the business or weather report pages of your newspaper), and averages (such as the usual Grade Point Average, or GPA, you had in school). These tools serve to describe actual observed data, which often can be quite large (e.g., every single grade you got in every single class you took in school), in a simple and summarized form (See Gonzalez-Marquez, Becker & Cutting this volume, for further details.)

Inferential statistics, on the other hand, was developed during the 20th century with the goal of providing tools for generalizing conclusions beyond actual observations. The idea is to draw *inferences* about an entire population from the observed sample. These are the tools that allow you to make conclusions, about, say "English speaking children learning irregular plurals," based on only a few observations of actual children (i.e., the fixed number of individuals constituting your sample, which is necessarily more limited than the entire population of English speaking children around the world). As we will see, making these kinds of generalizations – from a small sample to a huge general population – implies risks, errors, and making decisions about chance events. The goal of inferential statistics then is not to provide absolute certainty, but to provide robust tools to evaluate the likelihood (or unlikelihood) of the generalizations you would like to make. Later we will see how this is this achieved.

4. Variables and experiments

Statistical analyses have been conceived to be performed on collections of *numerical* data, where numbers are supposed to represent, in some meaningful way, the properties or attributes under investigation. Needless to say, if your observations are based only on verbal notes, sketches, interviews, and so on, you will not be able to perform statistical analyses directly on such raw material. In order to use statistics you will have to transform your raw data, via a meaningful procedure, into numbers. Sometimes, of course, that is not possible, or it is too pushy to do so, or it just does not make any sense to do it! In those cases, statistical analyses may simply not be the appropriate tools for your analyses (and you should be aware of that). However, when the properties or attributes under investigation (e.g., "degree of difficulty of a task," "College GPA average," or "the time it takes to answer a question") can take on different values for different observations, they are technically called *variables*.

Variables are extremely important in the context of experimental design, because they are at the core of the process of investigating *cause-effect relationships*. As it is well-known, one of the main goals of scientific research is to describe, to explain, and to predict phenomena (i.e., effects) in terms of their causes. The main way this is usually done is by carefully designing *experiments* where the investigator studies some specific variables, called *independent variables*, in order to determine their effect on the so-called *dependent variables*. The dependent variables are the ones that are measured by the investigator, which are expected to be affected by the independent variables. Independent variables are sometimes *manipulated variables* (when the experimenter directly manipulates the levels of the variable to be studied, like deciding the dose of a particular drug she will be giving to her participants) and sometimes they are *classifying variables* (when the experimenter categorizes the individuals according to some relevant criterion, for example, sex or medical condition).

Variables are also important in the context of *correlational studies*. In this case the investigators' goal is not to systematically manipulate an independent variable in order to observe its effect on a dependent variable (thus establishing a cause-effect relation) but rather to examine two or more variables in order to study the strength of their relationship. Correlational studies provide useful tools for investigating patterns and degrees of relationships among many variables, but they do not say much about causation. Rather, they say what is related with what, and to what degree the relationship is strong (or weak). When the relationship between two variables is strong it is possible to make predictions of values in one variable in terms of the other with an appropriate level of accuracy. A correlational study can determine, for example, that in general height and weight are two variables that are correlated: Very short people tend to weigh less than very tall people, and vice versa. The relation, of course, is not perfect. There are people who are short but with strong muscles and thick bones, and are therefore heavy, and also tall thin people who are light. But overall, if you observe hundreds or thousands of people, you will conclude that height and weight are quite related. There are specific statistical tools than can be used to establish precisely (in numerical terms) the strength and degree of relationship between variables. What is important to keep in mind is that the fact of establishing a strong relationship between two variables does not mean that we can directly make assertions about which one *causes* which one. We may establish that height and weight are very much related, but not because of this can we say, for instance, that "variations in height cause variations in weight" or that "variations in weight cause variations in height." We can say that height and weight are related (or very related, or even very strongly related) but we can not affirm which causes the other. Correlational studies are useful, however, to identify in a study with many variables with unknown behavior, which variables relate with which ones. Once the related variables are identified, the investigator may proceed to further study if a cause-effect relationship is suspected. But, in order to answer to that question, she will need to work within the framework of the experimental approach. For the rest of the chapter we will focus on the experimental dimension of empirical studies.

In order to better understand the concept of experimental design, consider a researcher who suspects that a drug G affects peoples' performance in remembering two-syllable nonsensical words. This suspicion, which usually is theory-driven and is technically called a *research hypothesis*, states that variations in the performance (dependent variable) are expected to be produced, or explained by variations in the dosage of drug G (independent variable). In this example the dependent variable could be defined operationally as "number of words correctly remembered from a list shown 15 minutes prior to the evaluation." In order to investigate her hypothesis empirically, our researcher will design an experiment, which should allow her to evaluate whether or not different doses of drug G produce changes in the performance of remembering those words. Then, she will collect a sample of people (hopefully selected randomly), and divide it into, say, four categories A, B, C, D. She will give to participants in each of the 4 groups, a specific dose of the drug: for instance, 10mg to those in group B, 40mg to those in group C, 100mg to those in group D, and 0mg to those in group A (the four dosage groups are technically called the *levels of the independent variable*, and they are usually determined by relevant information about the domain of investigation found in the literature). Then, through a carefully and systematically controlled procedure, she will show them a long list of two-syllable nonsensical words, and 15 minutes later, she will measure the individuals' performance (number of words remembered correctly) under the different dosage conditions (levels of the independent variable). She will then be ready to evaluate, using the tools of inferential statistics, whether the data support her hypothesis. (In reality, things are, of course, more complicated than this and get technical rather quickly, but for the purpose of this brief chapter we will leave it as it is.)

In this example, Group A plays a very interesting and important role. Why give 0mg to people in that group? Why bother? Why not simply eliminate that group if we are not giving them any drugs at all? Well, here is where we can see one of the main ideas behind an experimental design: *experimental control*. What the experimenter wants is to systematically study the possible cause-effect relationship between the independent and dependent variables. In our example, she wants to investigate how variations in drug G dosage (independent variable) affect remembering nonsensical words (dependent variable). But she wants to do this by isolating the effect of the independent variable while keeping everything else (or as much as possible) constant, thus neutralizing possible effects from extraneous variables. For instance, it could be the case that simply the testing situation, such as the list of words used for the study, or the environmental conditions in the room used for the testing, or the background noise of the computer's fan, or the psychological effect of knowing that one is under the "effect of a drug," seriously affect peoples' performance in an unexpected way, and this, independently of the actual active chemical ingredient of the drugs. Well, we would only be able to control the potential influence of such extraneous variables if we observe individuals in a group going through the full experimental procedure with one crucial *exception*: they do not actually get any active chemical ingredient believed to affect performance (i.e., absence of the attribute defined by the independent variable: 0mg of the drug). This means that the individuals in this group, known as the *control group*, should also get a pill, like everybody else in the study, except that, unbeknownst to them, the pill will not have the known active ingredient.

When studies like the one above are done properly, we refer to them as *experiments*, that is, controlled situations in which one (or more) independent variables (manipulated or classifying ones) are systematically studied to observe the effects on the dependent variable. The experiments can, of course, get more complicated. Our researcher, for instance, could be interested in studying the effects of more than one independent variable on the dependent variable. Other than the dosage of drug G, she could be interested in investigating also the influence of the time elapsed after showing the list of nonsensical words. This could be done through the manipulation of the variable "time elapsed," by measuring performance not only after 15 minutes of showing the list, but also at, say, 5 minutes, and 30 minutes, which would give a better sense of how time may affect performance. Or perhaps she could be interested in studying the influence of the number of vowels present in the nonsensical words (thus manipulating the number of such vowels), or even in how gender may affect the dependent variable (in this case it would not be a manipulated independent variable, but a classifying one), and so on. The more independent variables you include, the more detailed your analysis of how the dependent variable is affected, but also the more complicated the statistical tools get. So as a good researcher, you want to maximize the outcome of describing, explaining, and making predictions about your dependent variable, and to minimize unnecessary complications by systematically and rigorously studying relevant variables. The choice of what variables, how many of them, and what levels of each of them you need to incorporate in your experiment is often not a simple one. A beautiful experiment usually is one that explains a lot with only a few optimally chosen manipulations, the true mark of an experienced experimenter.

5. Measuring and measurement scales

The process of assigning numbers to attributes or characteristics according to a set of rules is called *measurement*. Measurement thus results in data collected on variables. In the everyday sense of "measuring," you measure, say, the temperature of a room by bringing in a thermometer and then reading the displayed value. But in the context of statistical analysis you are also measuring when, via a set of rules, you assign, say, the number "1" to "girls" and "0" to "boys" with respect to the variable "sex", or "1" to "low performance," "2" to middle performance," and "3" to high performance" with respect to performance in remembering nonsensical words from a list. Depending on the properties of your measuring procedure, different *measurements scales* are defined. It is very important that we know what measurement scale we are dealing with as they define the kind of statistical analysis we will be allowed to perform on our data. Simpler measurement scales only allow the use of simple statistical techniques, and more sophisticated measurement scales allow for more complex and richer techniques. The following are the four main measurement scales one needs to keep in mind.

5.1 Nominal scale

This is the simplest measurement scale, which classifies observations into mutually exclusive categories. This classification system assigns an integer to categories defined by differences in *kind*, not differences in degree or amount. For example, we use nominal measurements every time we classify people as either left- or right-handed and by assigning them the number 1 or 2, respectively. Or when we assign a number to individuals depending on their countries of citizenship: 1 for "USA", 2 for "Canada", 3 for "Mexico", and 4 for "other". In these cases, numbers act as mere labels for distinguishing the categories, lacking the most fundamental arithmetic properties such as order (e.g., greater than relationships) and operability (e.g., addition, multiplication, etc.). With nominal measurements, not only it is arbitrary to assign 1 to USA, and 3 to Mexico (it could be the other way around), but also it does not make any sense to say that the "3" of Mexico is greater than the "1" of the USA, or that the "2" of Canada is one unit away from the category USA. Because of the lack of arithmetic properties, the data obtained by nominal measurements are referred to as *qualitative data*. As we will see later, the statistical analysis of this kind of data requires special tools.

5.2 Ordinal measurement

This measurement scale is like the previous one in the sense that it provides a mutually exclusive classification system. It adds, however, a very important feature: order. Categories in this case, not only are defined by differences in kind, but also by differences in *degree* (i.e., in terms of greater-than or lesser-than relationships). For example, we use an ordinal measurement when we classify people's socioeconomic status as "low", "middle", or "high", assigning them the numbers 1, 2, and 3, respectively. Unlike in the nominal case, in ordinal measurements it does make sense to say that the value "3" is greater than "1", because the category assigning the value "3" denotes higher socio-economic status than the category assigning the value "1". This measurement scale is widely used in surveys, as well as in psychological and sociological studies, where, for instance, researchers classify people's replies to various questions as follows: 1="strongly disagree"; 2="disagree"; 3="neutral"; 4="agree"; 5="strongly "agree".

Numbers, in ordinal measurements, have the fundamental property of order, and as such they can be conceived as values placed along a dimension holding greater-than relationships. These numbers however, lack arithmetic operability. Because of this reason, the data obtained with ordinal measurements are also called qualitative data. Usually the statistical analyses for analyzing these data require the same tools as those used for nominal measurements.

5.3 Interval measurements

Interval measurements add a very important property to the ones mentioned above. They not only reflect differences in degree (like ordinal scales), but also have the fundamental property of preserving differences at equal intervals. The classic example of interval measurement is temperature as measured in Fahrenheit or Celcius degrees. Again, values can be conceived of as ordered along a one-dimensional line, but now they hold the extra property of preservation of equal intervals between units. Anywhere along the Fahrenheit scale, equal differences in value readings always mean equal increases in the amount of the attribute, which in the case of temperature corresponds to the amount of heat or molecular motion. In these measurements, arithmetic differences between numbers characterize the differences in magnitude of the measured attribute. This property guarantees that the difference in amount of heat between, say, 46°F and 47°F, is the same as the one between 87°F and 88°F (i.e., 1°F). Numbers used in this scale then have some operational properties such as addition and subtraction. Because of these arithmetic properties, the data are referred to as *quantitative data*. One important feature of these measurements is that the value zero (e.g., 0°F) is chosen arbitrarily: The value "zero" does not mean absence of the attribute being measured. In the case of the Fahrenheit scale, 0°F does not mean that there is *absence* of molecular motion, i.e. heat.

5.4 Ratio measurement

This is the most sophisticated measurement scale. It gives the possibility of comparing two observations, not only in terms of their difference as in the interval measurement (i.e., one exceeding the other by a certain amount), but also in terms of *ratios* (e.g., how many *times* more is one observation greater than the other). Ratios and times require numbers to be operational under multiplication and division as well, and this is exactly what this measurement scale provides. And in order for the idea of "times" to make any sense at all, the ratio measurement defines a non-arbitrary *true zero*, which reflects the true *absence* of the measured attribute. For example, measuring peoples' height or weight makes use of ratio measurements, where, because of a true zero (lack of height or weight, respectively) the measurement reads the *total amount* of the attribute. In these cases then, it does make sense to say that a father's height is *twice* that of his daughter's, or that the weight of a 150-pound person is *half* that of a 300-pound person. In the context of cognitive linguistics, examples of ratio measurements are "reaction time" or "numbers of words remembered correctly," where it is possible to affirm, for instance, that participant No 10 responded three times as fast as participant No 12 did, or that participant No 20 remembered only 25% of the words remembered by participant No 35.

These kinds of comparisons provide rich information that can be exploited statistically, but if there is no true-zero scale they cannot be made meaningfully. For instance, in the case of temperature mentioned earlier, it is not appropriate to say that 90°F is three times as hot as 30°F. One could say, of course, that the difference between the two is 60°F, but since the Fahrenheit scale lacks a true zero, no reference to ratios or times should be made. Because of the arithmetic properties of ratio measurements (i.e., order, addition, subtraction, multiplication, division), ratio data are *quantitative data* allowing for a wide range of statistical tools.

As we said earlier, when we are measuring, it is very important to know with what kind of measurement scale we are dealing with, because the type of scale will determine what kind of inferential statistical tests we will be allowed to use. Table 1 summarize the properties discussed above.

6. Samples and populations

A population is the collection that includes *all* members of a certain specified group. For example, all residents of California constitute a population. So do all patients following a specific pharmacological therapy at a given time. As we saw earlier, when researchers in social sciences are carrying out their investigations, they rarely have access to populations. Studying entire populations is often extremely expensive and impractical, and sometimes simply impossible. To make things tractable, researchers work with subsets of these populations: samples. Carrying investigations with a sample can provide, in an efficient way, important information about the population as a whole.

When samples are measured researchers obtain data, which is summarized with numbers that describe some characteristic of the sample called *statistics* (this notion of

Table 1. Measurement scales

Scale	Properties	Observations reflect differences in	Examples	Type of data
Nominal	Classification	Kind	Sex; major in college; native language; ethnic background	Qualitative
Ordinal	Classification Order	Degree	Developmental stages; academic letter grade;	Qualitative
Interval	Classification Order Equal intervals	amount	Fahrenheit temperature; Grade Point Average*; IQ score*	Quantitative
Ratio	Classification Order Equal intervals True zero	total amount and ratio	Reaction time; number of words remembered; height; income	Quantitative

* Technically, these are ordinal measurements, but they have been built in such a way that they can be considered interval measurements.

"statistics", by the way, should not be confused with the one used in the terms "inferential statistics" or "descriptive statistics"). An example of a statistic is the sample *mean*, which is a type of average, a measure of central tendency of the sample. Another example is the *standard deviation*, which is a measure of variability of the sample (see Gonzalez-Marquez, Becker & Cutting this volume, for details on how to calculate these statistics). A statistic is thus a descriptive measure of a sample (e.g., the mean height of a sample of undergraduate students you take from your college). On the other hand, a measure of a characteristic of a population is called a *parameter* (e.g., the mean height of all undergaduate students in California). To distinguish between descriptive measures of samples and populations, two different kinds of symbols are used: Roman letters are used to denote statistics (sample measures) and Greek letters are used to denote parameters (population measures). For example, the symbol for sample mean is $\bar{X}$ (called "X bar") and the symbol for the population mean is μ (pronounced "mew").

Statistics, which are numbers based on *the* only data researchers actually have, are used to make inferences about *parameters* of the population to which the sample belongs. But because every time a sample is taken from the general population there is a risk of not matching exactly all the features of the population (i.e., the sample is by definition more limited than the real population, and therefore it necessarily provides more limited information), there is always some degree of uncertainty and error involved in the inferences about the population. Luckily, there are many techniques for sampling selection that minimize these potential errors, which, among others, determine the appropriate size of the needed sample, and the way in which participants should be randomly picked.

The point is that we never have *absolute* certainty about the generalizations we make regarding a whole population based on the observation of a limited sample. Those inferences will involve risks and chance events. We have to ask ourselves questions such as, How likely is it that the sample we are observing actually corresponds to the population that is supposed to characterize? Out of the many possible samples we can take from the population, some samples will look like the real population, others not so much, and others may actually be very different from the population. For example, imagine that we want to investigate what the average height of undergraduate students is in all US college campuses. Since we do not have time or resources to measure every single student in the country, we take a sample of, say, 1000 students. Theoretically, we know that a good sample would be one that appropriately characterizes the "real" population, where the sample's statistics match the population's parameters (e.g., the sample mean corresponds to the population mean μ, and the standard deviation of the sample is equal to the one in the population which is designated by the Greek letter σ). But what are the chances that we get such a sample? How about if due to some unknown circumstances we mostly get extremely tall people in our sample (e.g., if we happen to pick the student-athletes that are members of the basketball team)? Or perhaps, we get lots of very short people? In those unlikely cases our sample's statistics will not match the populations parameters. In the former case, the sample mean would be considerably greater than the population mean, and in the latter it would be considerably smaller than the population mean. Here is where evaluating in a precise way the likelihood of getting such rare samples becomes essential. This idea is at the heart of hypothesis testing.

7. Probabilities and the logic of hypothesis testing

In order to understand the logic of hypothesis testing, let us go back to our example involving the researcher who wanted to study the effects of drug G on people's performance in remembering nonsensical words. Her general hypothesis was:

> General Hypothesis:
> Drug G dosages affect people's performance in remembering two-syllable nonsensical words.

Recall that in an experimental design, what the researcher wants is to investigate relevant variables in order to establish cause-effect relationships between independent variables (often by directly manipulating them) and dependent variables. In order to keep the explanation of hypothesis testing simple, let us imagine that the researcher has decided to compare the performance of a group of participants who took drug G with the performance of participants who did not (control group). In this case she will be testing the hypothesis that a specific dosage of drug G (independent variable), say 40 mg, affects the performance in remembering two syllable nonsensical words, measured 15 minutes after they had been shown (dependent variable). Her research hypothesis (which is more specific than the general hypothesis) is then:

> Research Hypothesis:
> A dosage of 40mg of drug G affects (increases or decreases) the performance of people in remembering two-syllable nonsensical words 15 minutes after presentation.

Let us now say that she randomly selects a sample of 20 participants (who will be supposed to characterize the vast population of "people" the hypothesis refers to. Keep in mind that the hypothesis does not say that the drug affects the performance of just these 20 participants. Rather it says that it affects the performance of *people* in general). Then she randomly assigns half of the participants to each group: 10 to group A (control group), and 10 to the experimental group (B) who will be given a dose of 40mg. After showing them the list of two syllable nonsensical words in strictly controlled and well specified conditions (e.g., procedure done individually, in a special room in order to avoid distractions, etc.) she measures the number of words remembered by participants in each group and obtains the following results: on average participants in the control group remembered 14.5 words, against an average of only 9 words in the experimental group.

At this point, the crucial question for the researcher is the following:

> Can we simply say that the research hypothesis is supoprted? That drug G does appear to affect (decrease, according to the results) *people's* performance in remembering nonsensical words?

Well, the answer is no. She can certainly say, from the point of view of descriptive statistics, that the mean number of words remembered by *participants* in the experimental group (9 words) is smaller than the mean of those in the control group (14.5 words). But because her goal is to make inferences about the population of "people" in general based on her limited sample, she has to evaluate the likelihood that the observed difference between the sample means is actually due to her experimental manipulation of drug G and not simply due to chance. She will have to show that her results are a *very rare outcome* if obtained by chance, and therefore, that they can safely be attributed to the influence of drug G.

In inferential statistics, the way of reasoning is roughly the following: Any observed result (like the difference in the average of recalled nonsensical words) is attributed to chance unless you have support to the contrary, that is, that the result is in fact a very rare outcome that should not be attributed to mere chance. Inferential statistics provides many powerful tools for evaluating these chances and determining the *probability*[3] of the occurrence of results. In our example, what our researcher needs to do is to run a statistical test to determine whether the difference she observed in the averages can be attributed to chance or not.

When running a statistical test, the researcher will have to specify even further her research hypothesis and state it in formal terms (called *statistical hypotheses*). Because the logic underlying inferential statistics says that observed results are attributed to chance until shown otherwise, the hypothesis is stated statistically in two steps. First, the *null hypothesis* (H_0) is stated, which formally expresses (usually) what the researcher would like to disprove. Then the *alternative hypothesis* (H_1) is stated, which formally expresses

---

3. The probability of an event A, which is denoted as $P(A)$, is expressed by a ratio: the number of outcomes divided by the number of possible outcomes. The way in which this ratio is defined, determines that a probability value is always greater or equal to zero (impossibility) and smaller or equal to one (absolute certainty). For instance, the probability that after rolling a die you get the number 2 is 1/6, since "2" is the only outcome, out of a total of six possible outcomes, that satisfies the condition expressed by the event. Similarly, the probability that if you roll the dice twice you obtain an added result of 9 is 4/36, which is equal to 1/9, because there are four outcomes out of thirty-six possible outcomes, that satisfy the condition of adding up to 9. These four outcomes occur when the first die is 3 and the second is 6, when they are 6 and 3, when they are 4 and 5, and 5 and 4, respectively.

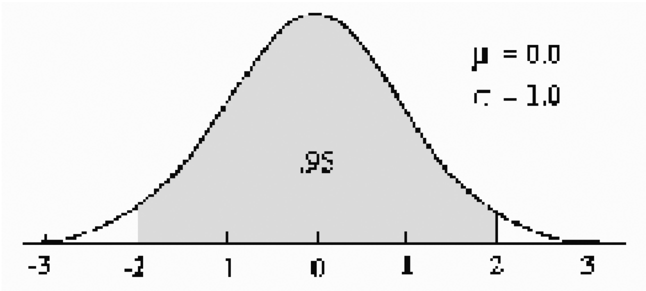


Figure 1. The normal distribution

the researcher's belief, the contrary of what H_0 states. Since at this level concepts have been characterized in terms of numbers and formal statements, H_0 is the formal negation of H_1. (i.e., usually it expresses the contrary of what the researcher wants to find evidence in favor of). In the drug G example, these *statistical hypotheses* would be expressed as follows:

Null hypothesis:
H_0: $\mu_A = \mu_B$

Alternative hypothesis:
H_1: $\mu_A \neq \mu_B$

The alternative hypothesis states that the population mean to which the control group belongs (μ_A) differs from the population mean to which the experimental group receiving the dose of drug G belongs (μ_B). The null hypothesis states the contrary: that H_1 is not true, that is, μ_A is not equal to μ_B. Again, remember that statistical hypotheses are stated this way because, as we said before, the logic of inferential statistics is to attribute any observed results to chance (H_0), until you can show the contrary (H_1).

In order to support the contrary possibility one has to show that the result is a very rare outcome if attributed only to chance. Statisticians can calculate exactly how rare is the result of an experiment performed on samples, by computing the probability of its occurrence with the help of theoretical distributions. An important distribution is the so-called normal curve, which is a theoretical curve noted for its symmetrical bell-shaped form, peaking (i.e., having its highest relative frequency) at a point midway along the horizontal axis (see Figure 1).

The normal curve is a precious tool for modeling the distributions of many phenomena in nature. One such phenomenon, for instance, is the distribution of height in a population. Figure 2 displays the distribution of heights of a randomly selected sample of 500 people. Height is shown as a variable on the horizontal axis, and proportion of occurrence on the vertical axis. Intuitively, we can see that most people's height falls around the value in the middle of the horizontal axis (i.e., at that middle point on *x* the proportion *y* is the highest, namely at 68 inches), that fewer people are either very short or very tall (e.g., if you look either to the left or to the right of the middle point, the proportions are smaller), and that even fewer are either extremely short or extremely tall (the proportions towards the extremes get smaller and smaller). If we get bigger and bigger samples,

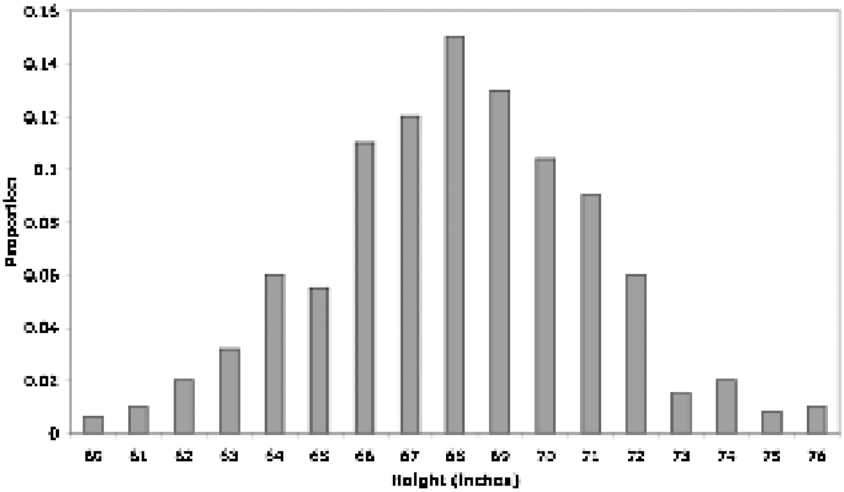


Figure 2. Distribution of heights of 500 people

our height data will progressively become more and more accurate in characterizing the distribution of heights in the population, and the resulting graph will get closer and closer to the theoretical normal curve.

The actual normal distribution, which as such is never reached, is a theoretical tool that has many important properties. One such property is that the theoretical middle point corresponds to the population mean, which other than being also the most frequent value, is the one that divides exactly the entire distribution into two halves (the normal curve has, of course, many more important mathematical properties which we cannot analyze here). For convenience, often we need to work with a *standardized* form of the normal curve, which, for simplicity, has been set with a mean of 0 ($\mu = 0$) and a standard deviation (which indicates the degree of variability of the population, and it is denoted by σ) of 1 ($\sigma = 1$). As a consequence, in the standardized form values below the population mean (0) are characterized in terms of negative numbers and those above the mean in terms of positive numbers. Finally, another important property of the normal curve is that it is always the case that 68% of all possible observations lie within 1 standard deviation from the mean, and approximately 95% of them lie within 2 standard deviation units (see Figure 1).

Keeping our drug G experiment in mind, the move now is to try to determine how rare is the difference of the observed sample means obtained by the researcher (a sample mean of 14.5 words for the control group vs. 9 words for the experimental group). The logic then goes as follows. If drug G really does not produce any effect at all, then whatever difference we may obtain can be attributed to chance. In such a case, we would expect the difference between the number of words remembered by the participants under the innocuous drug G and the number of words remembered by the participants in the control group, to be 0, or close to 0. So, if we think of the distribution we would obtain if we plotted the differences between the sample means (control minus experimental) of all possible pairs of random samples of 10 individuals (which is the size our researcher is working with) then we would obtain a theoretical normal distribution, technically called the *sampling distribution of sample means differences*. In this distribution, a difference of 0 between the sample means would be considered a very common outcome because it has the highest relative frequency. A positive but small difference between the two sample means would still be considered a relatively common outcome. So it would a small negative difference (where the sample mean of the experimental group is higher than the one of the control group). But a huge positive difference between the two sample means would indicate an extremely rare outcome if it was only attributed to chance (imagine the sample mean of the control group being, say, 100 words, and the sample mean of the group under drug G being only 2 words). If we depicted such difference (the standardize valve of 100 – 2 = 98) relative to its sampling distribution of sampling mean differences, the value would be located at the extreme right of the normal distribution whose relative frequency (*y* value) would be extremely small (i.e., highly unlikely to occur). Because the logic of inferential statistics is to attribute any result to chance until shown otherwise, statisticians have decided that in order to be extremely cautious, they would give the status of "rare outcome" only to those that are really rare, and not just "relatively" rare or "quite" rare outcomes. In numbers that means that they have decided to give the status of "rare outcome" only to those cases that are among the "rarest" 5%. The other 95% of cases are thus simply considered "common outcomes" (sometimes the criterion is even stricter, calling "rare outcome" only the 1% most rare, or even the 0.1% most rare, that is, those outcomes that are so rare that if they are only attributed to chance they would occur in at most 1 in a thousand times).

When brought into the values of the horizontal axis of the normal distribution, the criteria of "rare outcome" (e.g., the extremes 5%, 1%, or 0.1%) determine specific *critical values* which will help decide whether the result the researcher is interested in evaluating is a rare outcome or not. Based on this important idea of critical values, statistical tests have been conceived to produce specific numbers that reflect – numerically – how rare or common the results are (the underlying sampling distribution not always is the normal distribution. There are many other kinds of distributions, but for the sake of simplicity we will assume, for the moment, that the underlying distribution is normal). These numbers are compared to the critical values in order to decide which of the statistical hypothesis to reject and which one to accept: the null hypothesis H_0 (which says that nothing extraordinary or unusual is happening with respect to some characteristic of the population) or the alternative hypothesis H_1, identified with the research hypothesis, which indicates that something extremely unlikely is happening that should not be attributed to mere chance. When the underlying sampling distribution is the normal distribution, the extreme 5% (that is, the extreme 2.5% on each side of the curve) is determined by the critical value ± 1.96. If the statistical test produces a number that is greater than 1.96 or smaller than –1.96, we are in a position to reject the null hypothesis at a 5% *level of significance* (the outcome would be considered rare). If the value falls between –1.96 and 1.96 we retain the null hypothesis at 5% level of significance (the outcome would be considered common; see Figure 3).

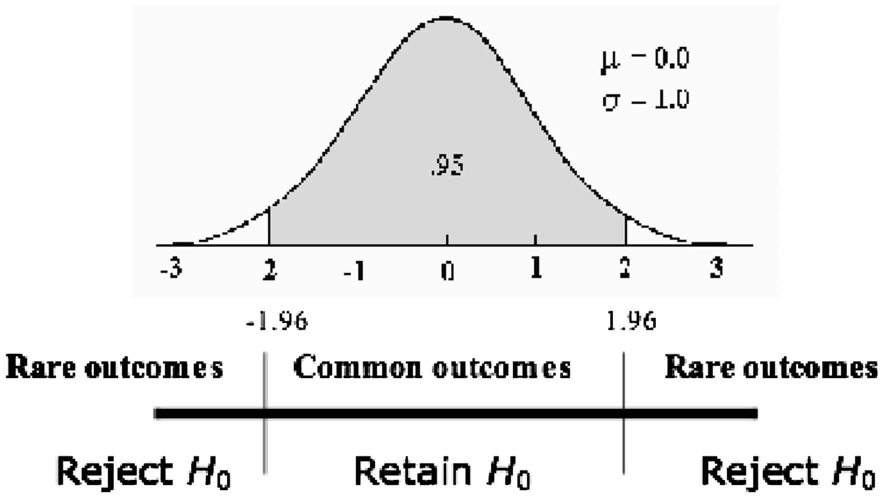


Figure 3. The standard normal distribution and the areas of retention and rejection of the null hypothesis

In our drug G example, if the value given by the appropriate statistical test (which we will see later) that evaluates the difference between the sample means of the groups (control minus experimental) is more extreme than the critical values (i.e., greater than the positive critical value on the right extreme of the normal curve, or smaller than the negative critical value on the extreme left side of the curve), then we would reject the null hypothesis and accept the alternative hypothesis. In such case we would conclude that the results observed in the experiment cannot be attributed to mere chance, and that they reflect the effect of the experimental manipulation (i.e., the effect of drug G). The particular test we would use in such case, which falls in the category of parametric inferential statistics, is called *t-test (for two independent samples*[4]*)*. There are, of course, many different inferential statistical tests that have been conceived for evaluating all sorts of data in different research designs (e.g., studies comparing three or more groups; studies with independent samples and studies with repeated measures; studies with one, two, or more independent variables; studies with one, two or more dependent variables; studies in which certain assumptions, such as the one of the normal distribution, cannot be made; studies whose data is not quantitative but qualitative; and so on).

4. Two independent samples refer to cases where observations in one sample are not paired with observations in the other one. A different version of the t-test for *matched samples* is used when the observations across the two samples are paired. This would occur, for instance, in a study of married couples that has the observation of each husband in one sample matched with the observation of the corresponding wife in the other sample. It would also occur when each pair of observations belong to the same participant measured under different conditions or at different times, for example before and after a particular treatment (this is also known as *repeated measures* design).

In the remainder of the chapter we will analyze some examples of statistical tests. We will go over the main concepts involved in two parametric tests, namely, the *t-test for two independent samples* (with a numerical example) and the *Analysis of Variance*, better known as *ANOVA*. And we will analyze a non-parametric test, the χ^2 (*Chi-Square*) *test* (with a numerical example). Let us now see the details of these statistical tests.

8. Parametric vs. non-parametric inferential statistics

The kind of statistical tests researchers are able to use is determined by several factors such as the type of measurement scale used to gather data (nominal, ordinal, interval, and ratio scales), and whether the technique has been conceived to test populations' parameters or not (there are other more technical criteria as well). When tests are about populations' parameters such as population means or population standard deviations, they are called *parametric* tests. When tests do not evaluate population parameters but focus on other features such as frequencies per categories (i.e., number of cases) they are called *non-parametric*. Usually statistical tests based on quantitative data are parametric, because, as we said earlier, when the data is obtained with interval and ratio measurement scales numbers have full arithmetic properties, which allow statistics such as sample means to be used for making inferences about the parameters of populations (e.g., population means). In other situations, when statistical tests do not focus on parameters, but in frequencies, we use non-parametric tests.

8.1 t-test, a parametric test

Let us go back to our example of whether drug G affects performance in remembering nonsensical words. The research hypothesis in that experiment is that a dosage of 40mg of drug G affects (increases or decreases) the performance of people in remembering two-syllable nonsensical words 15 minutes after presentation. Stated that way, the hypothesis does not specify in what way drug G "affects" performance, whether it increases or decreases performance. In such case, if the statistical test to be used makes use of the normal distribution (or the *t* distribution), the test is said to be *two-tailed* or *nondirectional* (i.e., the rejection areas are located at both tails of the distribution). But let us imagine that our researcher has more specific knowledge about drug G and that she is in a position to conceive a more specific experimental hypothesis:

> Experimental hypothesis:
> A dosage of 40mg of drug G affects (in fact decreases) the performance of people in remembering two-syllable nonsensical words 15 minutes after presentation.

Then she will require the use of a *one-tail* or *directional* test. By making the hypothesis that drug G *decreases* performance she is saying that she expects the difference between the means (control minus experimental) to be positive, therefore the rejection area is located in just one tail of the distribution (the positive side). The statistical hypotheses can now be specified as follows:

Null hypothesis:
H_0: $\mu_A - \mu_B \leq 0$

Alternative hypothesis:
H_1: $\mu_A - \mu_B > 0$

The difference of the population means (control minus experimental) is thus expected to be greater than zero (i.e., a dosage of 40mg of drug G decreases the performance of people in remembering two-syllable nonsensical words 15 minutes after presentation).

Let us now analyze the results obtained by our researcher. Here is the data:

Number of recalled words

Participant No	Control Group (A)	Experimental Group (B)
1	16	9
2	13	8
3	12	7
4	14	9
5	13	11
6	14	12
7	17	11
8	16	8
9	18	7
10	12	8
No of participants in each group (n):	10	10
Total sum of scores ($\sum X$):	145	90
Sample Mean ($\sum X/n$)		
Total sum/No of participants	145/10 = **14.5**	90/10 = **9**

In order to perform the t-test for two independent samples (which is the test recommended in this case, assuming that the underlying distributions are close enough to the normal distribution) we need to compute the following value:

$$t = \frac{(X_A - X_B) - (\mu_A - \mu_B)_{hyp}}{s_{X_A - X_B}}$$

But what does this formula mean? The value t will ultimately tell us whether the difference observed between the groups is rare or not. The greater the value of t, the more unlikely that the observed difference can be attributed to chance. If we look at the formula (ignoring the denominator $s_{X_A - X_B}$ for the moment), we can see that the value of t increases when the numerator increases, that is, when the difference between the sample means $(X_A - X_B)$ and the populations means $(\mu_A - \mu_B)_{hyp}$ increases. And this difference increases when the difference between the sample means $(X_A - X_B)$ is greater than the difference between $(\mu_A - \mu_B)_{hyp}$. Since the value of $(\mu_A - \mu_B)_{hyp}$ represents the hypothesized difference between population means (taken to be zero), then the crucial element that determines how big the value of t is going to be is the difference between the sample means $(X_A - X_B)$. If the sample means X_A and X_B are similar, then their difference is very small, and

t has a value close to zero. Such a small value of t, when compared to the critical values mentioned earlier (which ultimately determine what value of t the observed results are going to be called rare), will indicate that the little difference between the sample means can be attributed to chance, and therefore the null hypothesis H_0 is accepted (and the alternative hypothesis H_1 is rejected). But if X_A, the sample mean of the control group A, is much greater than X_B, the sample mean of the experimental group B, then that difference will be translated into a larger t value. If the t value is greater than the critical value that determines the limit of the rarest 5% (0.05 *level of significance*) then we can safely assume with a 95% of certainty that the observed difference between the sample means is a rare outcome if attributed to chance, and therefore we can reject the null hypothesis H_0 and accept the alternative hypothesis H_1.

In order to compute the actual value of t, we need to perform several calculations. Let us go step by step:

1) We know that the number of participants in each group is 10 ($n_A = 10$, and $n_B = 10$), and that
2) the total sum of scores per group is $\sum X_A = 145$ for the control group, and $\sum X_B = 90$ for the experimental group.

For reasons that we will not analyze here, we will also need some other calculations:

3) The sums of the squares of the scores per group:
$\sum X^2{}_A = 2143$ and $\sum X^2{}_B = 838$ (you can compute these numbers, by squaring each score and by adding them for each group).
4) An estimated measure of the variability of the sampling distribution of sample means differences (technically called *estimated standard error of* $s_{X_A - X_B}$, but whose details we will skip here. We will simply retain that, roughly speaking, this value expresses the average amount by which $(X_A - X_B)$ will deviate from its expected value by chance). Their computation is given by:
5)

$$s_A^2 = \frac{n_A(\sum X_A^2) - (\sum X_A)^2}{n_A(n_A - 1)} \qquad s_B^2 = \frac{n_B(\sum X_B^2) - (\sum X_B)^2}{n_B(n_B - 1)}$$

By substituting the numbers from our example into the formulae we obtain:

$$s_A^2 = \frac{10(2143) - (145)^2}{10(10-1)} \qquad s_B^2 = \frac{10(838) - (90)^2}{10(10-1)}$$

$$s_A^2 = \frac{21430 - 21025}{10(9)} \qquad s_B^2 = \frac{8380 - 8100}{10(9)}$$

$$s_A^2 = \frac{405}{90} = 4.5 \qquad s_B^2 = \frac{280}{90} = 3.1$$

6) With the value obtained in (4), we can now estimate the population variance with the pool variance estimate s_p^2, based on a combination of the two sample variances. Since n is the same in both groups ($n = 10$) this number will turn out to be an average of s_A^2 and s_B^2, namely 3.8.

7) We then compute the denominator of the t ratio, $s_{X_A - X_B}$:

$$s_{X_A - X_B} = \sqrt{[(s_p^2/n_A) + (s_p^2/n_B)]} = \sqrt{[(3.8/10) + (3.8/10)]} = \sqrt{[(0.38) + (0.38)]}$$

$$= \sqrt{(0.76)} = 0.87$$

Finally, we can compute our t ratio by substituting our numbers into the formula given above:

$$t = \frac{(X_A - X_B) - (\mu_A - \mu_B)_{hyp}}{s_{X_A - X_B}} = \frac{(14.5 - 9) - 0}{0.87} = 6.32$$

And we can compare our t ratio with the critical value that determines what difference between sample means are we going to consider a rare outcome. Critical values are usually found in special tables built after the corresponding distributions that, depending on the size of the samples (which determines what is called the number of *degrees of freedom* of the distribution) and the targeted level of significance (i.e., whether we want to call a rare outcome only those among the rarest 5%, 1%, or 0.1%), give you the critical value against which we have to compare our value in order to decide whether to reject or accept the null hypothesis H_0 (In this case, our t value is compared not with values from the normal distribution, but from a similar distribution, the t distribution[5]). In our example, after looking at the tables (with the appropriate of degrees of freedom, $10 - 1 = 9$ step which we will skip here) we see that the critical values for a one-tailed test and for a group sample size of $n = 10$, are 1.833 for a 5% level of significance, 2.821 for a 1%, and 4.297 for a 0.1% (notice that the values are increasingly bigger as we are more and more strict in deciding what are we going to call a "rare" event). Since the t ratio based on the drug G data is 6.32, which is greater than 4.297, the most strict critical value (0.1%), we decide to reject the null hypothesis at a .01% level of significance and accept the alternative hypothesis.

We are finally in a position to answer the original question: Does drug G decrease *people's* performance in remembering two-syllable nonsensical words? The experimental data obtained by our researcher allows us, via the t-test, to support the research hypothesis that said that drug G diminishes people's performance in remembering those words. The purely numerical result of the t-test can now be interpreted by saying that the difference between the number of two-syllable nonsensical words recalled by the participants in the two groups was big enough (control group being greater than the experimental group) that if it were to be attributed to chance it would be an extremely rare event occurring, at most, once every thousand times. Therefore, the fact that the performance of participants in the experimental group (B) who received drug G was lower than the one of participants in the control group (A) can safely be attributed to the performance-decreasing effect of drug G. This result thus provides empirical (experimental) evidence supporting the claim

5. The sampling distribution of t is almost equal to the standard normal distribution when the samples are higher than 30 observations (being equal to it at the theoretical case when there is an infinite number of observations). The t distribution has slightly inflated tails and its peak point is lower than the standard normal distribution. These features are more apparent when the samples are small.

that the decreased performance in the word recall task while taking drug G is very unlikely to be the result of chance.

8.2 Analysis of Variance (ANOVA), a parametric test

In the previous section we saw how the *t*-test is an effective tool for testing statistical hypotheses that compare the observed means of two samples. But what if our researcher, still interested in investigating her General hypothesis (i.e., drug G dosages affect people's performance in remembering two-syllable nonsensical words), wants to study the effect of the drug in a more subtle way by observing what happens when people are given different dosages of drug G, say, 20mg, 40mg, and 60mg? Her research hypothesis would now be:

> Research hypothesis:
> Dosages of 20mg, 40mg, and 60mg of drug G affect differently the performance of people in remembering two-syllable nonsensical words 15 minutes after presentation.

Our researcher needs to compare the performance of participants from not just two samples, but from several different samples (i.e., corresponding to different dosages of drug G). Assuming that a group A is the control group, and that groups B, C, and D are the groups whose participants receive 20mg, 40mg, and 60mg, respectively, then she needs to test the following statistical hypotheses:

> Null hypothesis:
> H_0: $\mu_A = \mu_B = \mu_C = \mu_D$
>
> Alternative hypothesis:
> H_1: $\sim(\mu_A = \mu_B = \mu_C = \mu_D)$

The null hypothesis H_0 states that the mean performances (remembering words) of the populations of people who have taken the specified dosages of drug G are the same. The alternative hypothesis H_1 says the contrary, i.e., that it is not the case that the mean performances are all equal. In order to test these hypotheses our researcher will need a statistical test designed to compare two or more samples.[6] One such test (a widely used one) is the *Analysis of Variance*, ANOVA. Because in this case there is only one independent variable involved (called factor), namely, the dosage of drug G, the analysis needed is a One-

6. The reader may be wondering why not use several *t*-tests to compare two observed means at a time (i.e., μ_A vs. μ_B; μ_A vs. μ_C; μ_A vs. μ_D; μ_B vs. μ_C; etc.) rather than using a new test. The reason is that the *t*-test has been conceived for comparing a *single pair* of observed means, not for doing multiple comparisons. Each time a statistical test is performed there is the probability of erring in rejecting a true null hypothesis. This is known as *Type I error*, and the probability associated with denoted by α, which corresponds to the level of significance chosen for rejecting the null hypothesis. The use of multiple *t*-tests increases exponentially (with the number of comparisons) the probability of Type I error beyond the value specified by the level of significance. Using one test comparing several means at a time (ANOVA) avoids that problem. Moreover, if the null hypothesis is rejected in ANOVA, there are related tests (e.g., Scheffé's test) that evaluate the observed difference between means for any pair of groups without having the cumulative probability of Type I error exceeding the specified level of significance.

way ANOVA (there are also more complex variations of ANOVA for repeated measures designs, for two factors, and so on).

The overall rationale of ANOVA is to consider the total amount of variability as coming from two sources – the variability *between groups* and the variability *within groups* – and to determine the relative size of them. The former corresponds to the variability among observations of participants who are in different groups (receiving different dosages of the drug), and the latter corresponds to the variability among the observations of participants who receive the same treatment. If there is any variation in the observations of participants belonging to the same group (i.e., receiving the same dosage of the drug), that variation can only be attributed to the effect of uncontrolled factors (*random error*). But if there is a difference between group means it can be attributed to the combination of random error (which, being produced by uncontrolled factors, is always present) *plus* the effect due to differences in the experimental treatment (*treatment effect*). In other words, the more the variability between groups exceeds the variability within groups, the higher the support for the alternative hypothesis (and the more the null hypothesis becomes suspect). This idea can be expressed arithmetically as a ratio, known as the *F ratio*:

$$F = \frac{\text{Variability between groups}}{\text{Variability within groups}}$$

This ratio is, in certain respects, similar to the t ratio analyzed in the previous section. As we saw there, the t ratio – used to test the null hypothesis involving two population means – corresponds to the observed difference between the two sample means divided by the estimated standard error (or pooled variance estimate s_p^2). The F ratio expresses a similar idea but this time it involves several sample means: the numerator characterizes the observed differences of all sample means (analogous to t's numerator, and measured as variability between groups), and the denominator characterizes the estimated error term (analogous to t's pooled variance estimate, and measured as a variability within groups).

The F ratio can then be used to test the null hypothesis (H_0: $\mu_A = \mu_B = \mu_C = \mu_D$) that all population means are equal, thus defining the F test[7]:

$$F = \frac{\text{Random error + treatment effect}}{\text{Random error}}$$

If the null hypothesis H_0 really is true, it would mean that there is no treatment effect due to the different dosages of drug G. If that is the case, the two estimates of variability (between and within groups) would simply reflect random error. In such case the contribution of the treatment effect (in the numerator) would be close to 0, and therefore the F ratio would be close to 1. But if the null hypothesis H_0 really is false, then there would be a treatment effect due to the different dosages. Although there would still be random error in both, the numerator and denominator of the ratio, this time the treatment effect (which generates variability between groups) would increase the value of the numerator resulting in an increase of the value of F. The larger the differences between observed group means

7. The F ratio has, like the t ratio, its own family of sampling distributions (defined by the degrees of freedom involved) to test the null hypothesis.

the larger the value of *F*, and the more likely that the null hypothesis will be rejected. In practice, of course – like in any statistical test – we never know whether the null hypothesis is true or false. So, we need to assume that the null hypothesis is true (which says that any observed differences are attributed to chance) and to examine the *F* value relative to its hypothesized sampling distribution (specified, like in the case of the *t*-test, for the corresponding degrees of freedom).[8]

So, how exactly is the value of the observed *F* calculated?[9] Mathematically, ANOVA builds on the notion of *Sum of Squares* (*SS*), which is the sum of squared deviations of the observations about their mean.[10] In other words, the *SS* is the technical way of expressing the idea of "positive (squared) amount of variability" which is required for characterizing the meaning of the *F* ratio. In ANOVA there are various *SS* terms corresponding to the various types of variability: $SS_{between}$ (for between groups), SS_{within} (for within groups), and SS_{total} for the total of these two. When the $SS_{between}$ and SS_{within} are divided by their corresponding *degrees of freedom* (the former determined by the number of groups, and the latter by the number of observations minus the number of groups), we obtained a sort of "average" amount of squared variability produced by both, between group differences and within group differences. These "averages" are variance estimates and are technically called *Mean Squares* (*MS*): $MS_{between}$ (for between groups), MS_{within} (for within groups). The *F* ratio described above is precisely the division of these two Mean Squares:

$$F = \frac{MS_{between}}{MS_{within}}$$

This value of *F* is then compared to the critical value associated to the hypothesized sampling distribution of *F* for specific degrees of freedom, which are determined by the number of observations and the number of groups, and for the specified level of significance. If based on her data, our researcher's observed *F* value is greater than the critical value, then the null hypothesis (that stated that all populations means were equal) is rejected, and she can conclude that different dosages of drug G differentially affect the performance in recalling nonsensical words.

---

8. The hypothesized sampling distribution of *F* is quite different from the normal distribution analyzed earlier. Unlike the standardized normal distribution and the *t* distribution, which are symmetrical relative to the axis defined by the mean (0) thus having negative values on one side and positive on the other, the *F* distribution only has positive values. This is due to the fact that, mathematically, the variabilities between groups and within groups are calculated using *sums of squares* (i.e., the sum of squared deviations of the observations about their mean), which are always positive. Because of this reason, the *F*-test is a nondirectional test.

9. The actual calculations for a complete analysis of variance are far too lengthy for the space allotted. Please see the statistical textbooks referenced here.

10. The differences are squared in order to avoid a problem inherited from a simple property of the mean: the sum of all the deviations of the observations about the mean equals zero. Squaring nonzero deviations always produces positive numbers, and therefore it eliminates the problem of having the deviations "canceling" each other out. A sum of squares is thus always greater or equal than zero.

8.3 χ^2 (Chi-Square), a non-parametric test

Now let us look at a statistical test that it is often used when the data is qualitative or obtained in categorical form. In order to illustrate this technique let us analyze a research example in the study of conceptual metaphor.

Cognitive linguists have analyzed many languages around the world noticing that in all of them, spatial language is recruited to talk about time. As in any conceptual metaphor, the theory says, the inferential structure of target domain concepts (time, in this case) is via a precise mapping drawn from the source domain (uni-dimensional space, in this case). After examining hundreds of English expressions involving time, Lakoff and Johnson (1980) identified two different metaphorical cases, Time Passing Is Motion of an Object (as in the expression *Christmas is coming*) and Time Passing Is Motion Over a Landscape (e.g., *we are approaching the end of the month*) (Lakoff 1993). The former model (the *Moving-Time* version) has a fixed canonical observer where times are seen as entities moving with respect to the observer (Figure 4), while the latter (the *Moving-Ego* version) sees times as fixed objects where the observer moves with respect to time (Figure 5).

Psychologists have since questioned whether there is any psychological reality in people's minds when they listen to, or utter, such metaphorical expressions (see, for example Murphy 1997). Is it the case that people actually operate cognitively with these conceptual metaphors? Or could it be the case, as some scholars have argued, that the temporal metaphorical expressions are simply "dead metaphors," that is, expressions that once had spatial content but that now have become separate temporal lexical items, no longer with

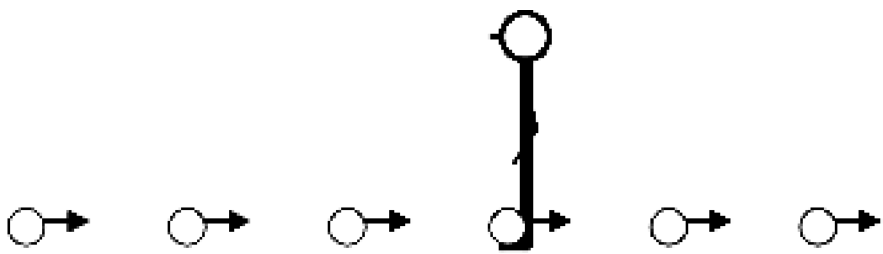


Figure 4. A graphic representation of the Time Passing Is Motion of an Object metaphor

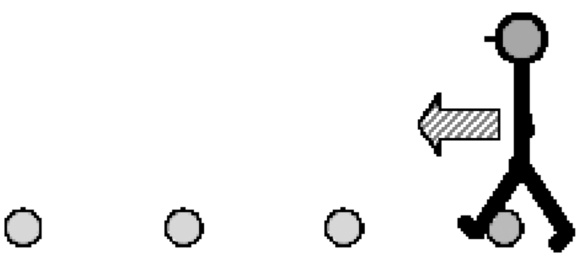


Figure 5. A graphic representation of the Time Passing Is Motion Over a Landscape metaphor

connections with space? Psychologists have tried to answer some of these questions using priming techniques, a well-known experimental paradigm in which participants are systematically biased via specific stimulation involving the source domain in order to evaluate whether they carry the corresponding inferences into the target domain. If by priming the source domain of the metaphor (spatial) one gets systematic variation in the inferences made in the target domain, then one could conclude that individuals do reason metaphorically, or otherwise they would not be sensitive to the priming. With this paradigm psychologists have gathered experimental evidence that indeed real or represented physical motion scenarios can prime different construals of time (Boroditsky 2000; Gentner, Imai, & Boroditsky 2002). For example, speakers who have just been moving (e.g. traveling on a plane or a train) or imagining self-motion are primed to give Moving-Ego rather than Moving-Time interpretations of metaphorical time phrases in English, such as "the Wednesday meeting was moved forward two days." This phrase can be interpreted according to either of two mappings, Ego-moving or Time-moving; if Time is seen as moving towards Ego, then earlier events are "ahead of" later ones, and the meeting is seen as rescheduled to an earlier time (Monday); while if Ego is moving through space, then farther future events are farther ahead relative to Ego, so moving the meeting ahead is seen as rescheduled to a later time (Friday). With a priming background of self-motion, respondents' interpretation is thus biased towards understanding forward as "to a later time" (i.e., via Moving-Ego metaphor).

But can we really assume that people replying "Monday" rather than "Friday" are making the inferences in terms of a temporal motion "towards Ego"? Perhaps in the case of "Monday" answers, what people are doing is to make inferences in terms of the reference point (RP) in a sequence where the presence of an Ego is completely irrelevant. In such case, we would have a more primitive mapping where *in front of the sequence is earlier and behind is later* (Figure 6), without reference to any Ego (which in the target domain would bring the moment "Now"). This more fundamental metaphor (called *Time-RP* as opposed to *Ego-RP*, by Núñez and Sweetser (2006), would model linguistic expressions such as *spring follows winter* or National Public Radio's announcement *twenty minutes ahead of one o'clock* (to refer to 12:40) which recruit dynamic language but where there is no reference to the present "now."[11] Núñez (1999) and Moore (2000) have already provided linguistic and theoretical analysis of such more elementary metaphorical case. But what can be said of the psychological reality of this Time-RP metaphor? What we would need now is not to add more linguistic examples to the list of expressions, but to gather empirical evidence of its *psychological reality*.

Figure 6. A graphic representation of the Time-RP metaphorical construal of time

11. An Ego, of course, could be added to this mapping (thus bringing a "Now"), resulting in the so-called Time Passing Is Motion of an Object (Figure 4). But such a case wouldn't be a primitive mapping but a more complex composite one.

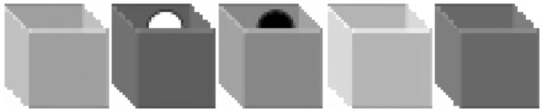


Figure 7. Example of the priming stimulus which was presented dynamically going from one side of the screen to the other

So, coming back to the ambiguous expression "the meeting was moved forward," if participants reason in terms of this metaphorical mapping, then "to move forward" would be seen as moving towards earlier moments in the sequence without having to refer to an Ego (i.e., now). This could be stated as a research hypothesis:

> Research hypothesis:
> The proportion of people responding "Monday" to the ambiguous sentence "Wednesday's meeting has been moved forward two days. When is the meeting?" will increase, if they are primed with a visual stimulus showing a simple sequence of cubes going from one side to the other of the computer screen.

In order to test such a hypothesis, in my lab we conducted an experiment in which the main goal was to prime the (spatial) source domain with an Ego-less sequence of cubes moving on the screen from one side to other (see Figure 7). The idea was to see whether participants would get prompted to give "Monday" replies just by looking to a simple sequence of moving cubes (with no reference whatsoever to an Ego). To avoid the bias that a left-to-right vs. right-to-left presentation would produce, the priming stimulation was done in a counterbalanced way. Participants in the control group observe only a static verb of the cubes.

134 students participated in the study, half of which were in the control group (no priming) and half in the experimental group (with priming). The statistical hypotheses are the following:

> Null hypothesis:
> H_0: $P_c = P_e$

Here P_c denotes the proportion of people in a non-primed population replying "Monday" rather than "Friday" to the above question (control), and P_e denotes the proportion of people answering "Monday" in a population primed with the visual stimulus (experimental). The null hypothesis says that the proportions are the same.

> Alternative hypothesis:
> H_1: $P_c \neq P_e$

As usual, the alternative hypothesis states that the null hypothesis is false, that is, that the proportions of the two populations are different.

The following are the results of the experiment. The table below shows the number of participants categorized according to their answers. The numbers indicate the frequency of occurrence for each cell. Because the data here have been measured with a nominal scale the data are considered qualitative (participants' have been categorized according to their responses, "Monday" or "Friday"). With this kind of data the statistical test we can use in order to answer which statistical hypothesis is correct is the *one-way* χ^2 (*Chi-Square*) *test*. This test is a non-parametric test because it does not deal with specific parameters of the population such as population means, but rather it deals with frequency of occurrence of categorical observations.

	Monday	**Friday**	Total
Priming (Exp.)	**42**	**25**	67
No priming (Control)	**26**	**41**	67
Total	68	66	134

At a first glance we can see that without any priming more participants replied "Friday" to our question. In our sample, about a 61% of the participants in the control group (non-primed) replied "Friday" (41 participants out of 67), and a 39% of them replied "Monday" (26 out 76 participants). This set of data goes in the same direction as the findings reported by Boroditsky (2000), where 54% of the participants responded "Friday" and 46% responded "Monday." If we look at the participants in the experimental group (primed), however, the proportion of "Monday" replies (42 out of 67, or 63%) is greater than the proportion of "Friday" replies (25 out of 67, or 37%). We now need to use the χ^2 *test* in order to determine whether this difference in proportions is a rare outcome if we attribute it to chance or not.

Like with the *t-test*, we need to calculate a value (χ^2) which we will compare with its hypothesized sampling distribution (which this time is different from the normal distribution. For details see any specialized text). The calculation is done with the following formula:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

where f_o denotes the observed frequency of "Monday" and "Friday" responses from participants in the experimental group, and f_e denotes the expected frequency for each category ("Monday" and "Friday") from participants in the experimental group. "Expected frequency", in this example, means that if participants do not go through any special treatment (such as the priming stimulation the participants in the control group were exposed to), then that is the frequency of "Monday" and "Friday" responses one would expect to encounter in the unprimed population. We can see that, as with the *t-test*, the value of χ^2 will increase if the numerator increases. And the numerator increases if the sum of the squared differences between the frequencies of the control group and the experimental group increases (The differences are squared, thus guaranteeing that the value will be always positive and therefore the amount of differences will cumulate only in positive terms). In short, we will get a big χ^2 if the differences between expected and observed frequencies are big, and it will be small, or close to zero, if those differences are small. The test will help us determine if the differences we found between expected and observed frequencies are significantly large or not. The following is the computation of χ^2 for the results of the above experiment:

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(42-26)^2}{26} + \frac{(25-41)^2}{41} = \frac{256}{26} + \frac{256}{41} = 16.09$$

When compared to the appropriate sampling distribution of χ^2 (with corresponding degrees of freedom), the obtained value $\chi^2 = 16.09$ is greater than the critical values at 5%, 1%, and even 0.1%, which are 3.84, 6.63, and 10.83, respectively.

So we can now reject, with a .01% level of significance, the null hypothesis H_0 (which says that the proportions of people replying "Monday" rather than "Friday" in both groups are similar), and accept the alternative hypothesis H_1: The proportions of "Monday" responses are indeed (very) different (being that proportion greater in the experimental group). And the difference is so big that if we were to attribute it to chance, it will occur at most one in a thousand times.

In terms of the research question, the interpretation of this result is that a visual stimulus composed of an Ego-less sequence of cubes, which intervenes at the level of the spatial source domain of the conceptual metaphor, primes people such that answers to the ambiguous temporal question interpreting "forward" to mean "earlier" in time, increase significantly. This result thus provides experimental evidence supporting the psychological reality of the primitive Time-RP metaphor depicted in Figure 6 (for a more complete account with similar experiments see Núñez, Motz, & Teuscher 2006).

9. Epilogue

We have briefly analyzed some statistical tests, two parametric tests (*t*-test and ANOVA), and one non-parametric (χ^2, Chi square). In the world of inferential statistics there are many more techniques available to the researcher, including several variations of the techniques we saw. Which test to use will depend on many factors, such as the nature of the research question(s), the kind of measurement scale used for gathering the data (quantitative or qualitative), the number of independent and dependent variables, the degree to which the data satisfy the assumptions that many of the statistical tests require (e.g., certain statistical tests require that the variability of the distributions to be compared is similar, what is known as *homoscedasticity*), the degree to which the underlying theoretical distributions satisfy the required assumptions (e.g., many techniques require the underlying distribution to be normal), and so on.

It is important to keep in mind that statistical tests are conceptual *tools*, rigorously and systematically constructed mathematically, but they are conceptual tools nonetheless. As such, they have been conceived for specific purposes and built on the basis of certain assumptions. When the assumptions are not satisfied, or when the research purpose is radically different from the one for which the test has been designed, proceed very carefully. You can always do "number-crunching" and perform all kinds of blind calculations in all sorts of ways. But, to what extend does it make sense to perform those calculations? And to what extend are you "allowed" to do so if your data do not meet the requirements specified by the underlying assumptions? With the number of easy-to-use statistical packages available in the software industry today, it is easy to get excited about performing (sometimes meaningless) calculations, loosing the conceptual big picture. The moral is that we should not get lost in the forest of data, tables and numbers. It is always wise to step back and ask ourselves why it makes sense to pick one test instead of another one. We need to think of the assumptions underlying the various techniques as well as the limitations and advantages they present. When doing empirical research supported by inferential statistics we should always keep the meaningful big picture clearly in our minds.

Finally, and in order to close this chapter, I would like to address the question of who has the last word where scientific questions in *empirical cognitive linguistics* are concerned: Does the last word belong to linguists, who are doing the linguistics? Or does it belong to psychologists or neuroscientists, who are doing the empirical studies? The way in which I presented the material of this chapter may give the impression (shared by a certain number of experimental psychologists) that cognitive linguistics is basically a theoretical enterprise that, beyond the linguistic realm can, at best, generate interesting hypotheses about cognitive phenomena, but that it is up to experimental psychologists to finally decide, via empirical studies, on the ultimate truth of those statements. Under this perspective, progress in cognitive linguistics is achieved through the production of consecutive steps in (1) theoretical cognitive linguistics, which serves to generate hypotheses, followed by (2) empirical observations in experimental psychology meant to test those hypotheses. Experimental psychology thus has the last word as far as empirical cognitive linguistics is concerned.

The position I want to defend here, however, is quite different from that. For specific questions, such as the "psychological reality" of some particular cognitive linguistic phenomena (e.g., the psychological reality of a given conceptual metaphor), the process may indeed follow those steps. First, cognitive linguists describe and analyze the phenomenon in linguistic terms, and then the psychologists run the experiments to find out whether the phenomenon has some psychological reality. This may be the case of how things work when the subject matter is about "testing psychological reality." Psychological reality is, after all, a psychological phenomenon. But when the question is about the nature of progress in the field of empirical cognitive linguistics, I do not see experimental psychology as having the last word. I see it as playing an intertwined role along with linguistic studies (as well as other close empirical fields such as neuroscience and anthropology). Results in experimental psychology feed back into developments in linguistics, and vice versa, results in linguistic analysis feed back into experimental psychology. Moreover, this latter feed back is not only in terms of "generating" hypotheses for empirical testing, but also in terms of specifying what experimental distinctions to make, what relevant variables to manipulate, or even how to interpret current or past experimental findings (process which is never "theory-free").

We can take the last statistical example (χ^2, Chi square) involving the spatial metaphors for time to illustrate the previous point. We saw that in such case, first there was the work done by linguists and philosophers identifying, classifying and characterizing hundreds of linguistic metaphorical expressions for time (e.g., Lakoff & Johnson 1980). Then these linguists modeled the inferential structure via "conceptual metaphors," (i.e., inference-preserving cross-domain mappings) taken to be cognitive mechanisms. This resulted in two theoretical constructs, that is, two main conceptual metaphors for time: Time Passing Is Motion of an Object and Time Passing Is Motion Over a Landscape (Lakoff 1993). Then came the psychologists who, via priming studies, investigated the psychological reality of these "conceptual metaphors" (Boroditsky 2000; Gentner, Imai, & Boroditsky 2002). Now, the question is: Is that the last word about the question of spatial construals of time? I do not think so. The taxonomy used by the psychologists (and the linguists before them) considered only two cases, both of which were based on "what moves relative to what." But work done in theoretical cognitive semantics (Núñez 1999), cross linguistic studies (Moore 2000), and in gesture-speech investigations (Núñez & Sweetser 2006) argued for the necessity to re-classify spatial metaphors for time, this time not based on "what moves relative to what" but on "what is the reference point" in question, independently of whether there is motion involved. This allowed for a *re-interpretation of the empirical results* obtained by psychologists, and also for the generation of new hypotheses to be tested empirically, such as the question of the psychological reality of the more elementary *Time-RP* conceptual metaphor. Progress in empirical cognitive linguistics in this case was not made by (1) posing a theoretical question (in cognitive linguistics) after finding patterns in linguistics analyses, then (2) by answering it through experimental methods in psychology, and then (3) moving to the next question leaving the answer in (2) untouched. Rather, it was done (or better, it is still being done as we read these lines) by going through the first two steps, but then feeding those two steps with new linguistically-driven distinctions, such as the notion of *Reference-Point*, which could not have come (in fact it did not come) from empirical results in experimental psychology alone. It was the fact that the notion of reference point (rather than relative motion) was fed into the process of understanding spatial construals of time that generated, both, new theoretical interpretations of existing empirical results, and new hypotheses about human cognition and its psychological reality to be tested empirically.

The process of gathering knowledge in cognitive linguistics is, of course, open for new developments not only in linguistics and psychology, but in neuroscience and other new neighboring fields, as well. It is through the ongoing process of mutual feeding that genuine knowledge gathering is perpetuated.

References

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75*, 1–28.

Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space > time metaphors. *Language and Cognitive Processes, 17,* 537–565.

Gibbs, R. (1996). Why many concepts are metaphorical. *Cognition*, *61*, 309–319.

Heiman, G. W. (2003). *Basic Statistics for the Behavioral Sciences*. Boston, MA: Houghton Muffin.

Helmstadter, G. C. (1970). *Research concepts in human behavior: Education, psychology, sociology*. New York: Appleton-Century Crofts.

Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied Statistics for the Behavioral Sciences*. Boston, MA: Houghton Muffin.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Methaphor and Thought* (pp. 202–251). Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Moore, K. E. (2000). Spatial experience and temporal metaphors in Wolof: Point of view, conceptual mapping, and linguistic practice. Unpublished doctoral dissertation, University of California, Berkeley.

Murphy, G. (1997). Reasons to doubt the present evidence for metaphoric representation. *Cognition*, *62*, 99–108.

Núñez, R. (1999). Could the future taste purple? *Journal of Consiousness Studies, 6*, 41–60.

Núñez, R., Motz, B., & Teuscher, U. (2006). Time after time: The psychological reality of the ego- and time-reference-point distinction in metaphorical construals of time. *Metaphor and Symbol*, *21*, 133–146.

Núñez, R. & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science*, *30*, 401–450.

Sabourin, M. (1988). Méthodes d'acquisitions de connaisances. In M. Robert (Ed.), *Fondements et étapes de la recherche scientifique en psychologie*. St-Hyacinthe, Quebec: Edisem.

Stigler, S. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Witte, R. & Witte, J. (2007). *Statistics* (8th ed.). Hoboken, NJ: John Wiley & Sons.

Multiple empirical approaches to a complex analysis of discourse

Linda R. Waugh, Bonnie Fonseca-Greber, Caroline Vickers and Betil Eröz

1. Introduction

Our purpose here is to exemplify the benefits of an integrated approach to the study of discourse that examines culturally contextualized examples of authentic language use with a rich, fine grained analysis derived from various empirical approaches. By discourse we mean the actual use of language for communication. Communication has many functions (see Jakobson 1960), including referential (e.g., talking about the outside world), emotive (e.g., expressing our emotions, attitudes, beliefs), conative/directive (e.g., getting others to do things, making sure they pay attention to what we say), phatic (e.g., making sure our interlocutor can hear us), metalinguistic (e.g., talking about language), poetic (e.g., aesthetic use), and interpersonal (e.g., creating, maintaining, changing interpersonal relationships – see Halliday 1978) and we need to study language in all of these functions if we're really to have an understanding of language-in-use. Moreover, discourse and its communicative functions are intertwined with linguistic, cognitive, social, cultural, historical, ideological and biological patterns, and none of these exist separately from the others. In other words, ours is an ecological perspective on language, which views language as embedded in a complex ecological system in which all facets are co-dependent. As a result, discourse, here, will be viewed as a process that continually shapes and is shaped by its ecological embeddedness. This means that discourse influences and is influenced by all the other facets of its ecological setting, including all those factors that shape and are shaped by human cognition. This viewpoint is shared by a number of researchers (e.g., Kramsch 2002; Mittelberg et al., this volume; Lantolf 2000; van Lier 2000, 2004; for a more explicit focus on discourse and cognition, see Edwards 1997 and Virtanen 2004; for an integration of cognitive and functional approaches to language structure, see Tomasello 1998, 2003).

Our focus here will be on discourse and what it tells us about all of these factors since we feel that this is an important empirical starting point. Here we will be studying the discourse of members of specific speech communities (also called discourse communities), a speech community being defined as people who "use, value or interpret language" in shared ways (Saville-Troike 1982/1989/2003: 15). We will focus on the spoken language, since that is universal to all human beings, and will deal with only some types (genres) of speech events. A speech event (also sometimes called a communication event – see Jakobson 1960; Hymes 1974) is a specific instance of language use (including a linguistic message, an utterance, a piece of discourse), which takes place between particular people (speaker and addressee=interlocutors), at a specific time and place, in a particular social and cultural context, using (a) particular language(s), with a specific means of communication (e.g., ordinary speech, telephone, TV, amplified sound, etc.). The speech events analyzed here are comprised of two or more participants in face-to-face dialogue with each other and engaged in ordinary conversation between family members and friends, discussion among students doing a project together, discussion in a classroom, etc. We decided on these speech event types because we wanted to focus on language use that you, the readers of this book, are familiar with and could therefore relate to. All of the participants in our speech events share a common language (French in the case of Waugh and Fonseca-Greber=W&F-G, English in the case of Vickers and of Eröz, though for some participants it is not their native language); and in many cases, they know each other more or less well and have a history of linguistic and social interaction with each other. In some cases, through repetition and shared goals, the interaction between the participants has evolved into a community of practice, "an aggregate of people who come together around some enterprise. United by this common enterprise, people come to develop shared ways of doing things, ways of talking, beliefs, values – in short, practices – as a function of their joint engagement in activity" (Eckert 2000: 35) – all of which practices are related to cognition and conceptual structures.

Most importantly, we are focusing on authentic use of language – language that is socio-culturally and cognitively defined as a form of talk (Goffman 1981) that is typical of the given speech community/community of practice (spontaneously produced and not based on artificial situations) – since we believe that this is the only way to get at what language is really like. In some cases of authentic language, the analyst is present in the speech event and may even participate in it – as is the case for some of the speech events we report on here – but in all cases the analyst is careful not to bias what the participants say and how they say it (or at least, we do our best not to bias it – you can never be sure about this, of course, but we all did our best not to be an intrusive presence). Typically, discourse analysts from many different empirical traditions collect examples of such authentic language use through audio- or video-taping (or both). These traditions include, for example, (anthropological and socio)linguistic approaches to discourse analysis (Labov & Waletzky 1967; Halliday & Hasan 1976; Givón 1979; Hopper 1982; Klein-Andreu 1983; Tannen 1984, 1989, 1993; Chafe & Nichols 1986; Hill & Irvine 1992; Chafe 1994; Bybee & Fleischman 1995; Fox 1996; Cheshire & Trudgill 1998; Du Bois 2003), conversation analysis (CA) (Goodwin & Heritage 1990; Duranti & Goodwin 1992; Schiffrin 1994: 232–281; Hutchby & Wooffitt 1998; Markee 2000), interactional sociolinguistics (Goffman 1981; Gumperz 1982, 2001; Brown & Levinson 1987; Tannen 1984, 1989; Schiffrin 1987, 1994: 97–136, 2001), and corpus linguistics (Stubbs 1996; Kennedy 1998; Biber, Conrad & Reppen 1998; Tognini-Bonelli 2001; Conrad 2002). By taping, the details of language behavior will not be lost (Goodwin & Heritage 1990: 289; Biber, Conrad & Reppen 1998; and Edwards & Lampert 1993) and can be listened to time and again. This body of data has been called a corpus by some of the theoretical approaches we use here (but not by all – indeed, some would be surprised at our use of the word corpus, but we feel that it captures nicely the fact that we're using a body of authentic data gathered systematically under rigorous conditions). Since there is a chapter in this volume on corpus-based research, we will not address further the issue of corpus-design, except to say that, while most emphasis in corpus linguistics has been on large corpora (sometimes consisting of millions of words), here we will focus on smaller corpora within well-defined speech communities in well-defined speech events. Our corpora are large enough for quantitative analysis, and small enough to allow us to concentrate on the fine details of linguistic interaction since it's in the details of everyday talk that language works – and we can then relate these details to both the specifics of the speech event and the larger community of which they are a part. In other words, we prefer these smaller corpora that we know well and were gathered for our own work specifically, to the sometimes quite anonymous corpora of the large data banks.

In order to carry out our fine-grained work, we all worked with transcripts (this is standard in corpus work). Now, transcripts are not completely objective, since they are inherently selective and interpretive, and, since there are several different transcription conventions associated with different types of research (e.g., CA, Ethnography of Communication, various researchers doing corpus linguistics – see the contributions to Edwards & Lampert (Eds.), 1993), we each had to decide on the transcription conventions that would be suitable to the analysis being carried out, including the theoretical perspectives and the research questions being addressed (see Ochs 1979; and Edwards 1993, 2001). Transcription conventions are definitely not one-size-fits-all.

The most important issue is that the transcript has to reflect what is actually said, not some cleaned-up version of what is said. Real speech is filled with unfinished sentences, partially pronounced words, incoherent utterances, people overlapping with each other, and so forth – and these are an important part of language use and have to be dealt with in some way (which is why transcripts are different from the standard written language, although they are done in writing). However, it is not enough just to have a transcript. Linguistic communication takes place in a context, and thus, in order to understand discourse, we need also to take into account the socio-cultural systems that have an effect on/are effected by language use, the historical contingencies that influence/are influenced by language use, the ideological structures that permeate the world in which language is used, etc. This means, for example, that the researcher needs to collect social and cultural/ethnographic information in order to better interpret what the participants are saying and what it means to them (as practiced, for example, in the ethnography of communication, see Coulthard 1977/1985; Saville-Troike 1982/1989/2003; Fasold 1990: 39–64, Schiffrin 1994: 137–189), and in sociolinguistics, see Labov 1972, 2001; Coupland & Jaworski 1997; Hudson 1980/1996; Fasold 1984, 1990). For the purposes of our work, all of us gathered information about the participants themselves (e.g., age, sex, education, social status, occupation, native language), the relationship of the participants to each other and their roles in the speech event (whether determined in advance or negotiated during the interaction). In addition, we identified the type of speech event involved: e.g., family meals in France and Switzerland, students meeting outside of class in America to discuss a project they are jointly working on, students in an English composition classroom in America, etc. Each of these types is defined by the particular society of which they are a part (France, Switzerland, U.S.), as well as the type of context (familial, academic, institutional); hence, it was crucial for us to have knowledge of their socio-cultural status as well as norms and expectations about language use in each type for a complex and rich analysis of the data. Moreover, each has definite cognitive demands that need to be understood in their ecological complexity.

Now, this type of information about speech events is not in conformity with the approach taken, in particular, by CA, which "employs inductive methods to search for recurring patterns across many cases" (Lazarton 2002:37) and in its more orthodox version claims, for example, that the analyst should know nothing about the participants, except for what can be discerned from the actual talk. However, we agree with Moerman (1998), who advocates combining CA and ethnography of communication into culturally contextualized conversation analysis (CCCA) and proposes close analysis of conversational data in combination with information about the speech community within which the conversation occurs in order to gain an understanding of the culturally specific meaning of the data for the participants. Sometimes, participants in a speech event will not disclose what's taken for granted, because it's so obvious to them on socio-cultural, conceptual, cognitive, historical grounds – and it's important for the analyst to have some sense of the taken-for-granted. At the same time, the analyst, Moerman says, "tries to limit the ingredients of interpretation, the components of meaning, to ones that are locally significant and locally occasioned" (Moerman 1988:7). These abductive interpretations (Peirce 1878/1992) by the analyst are akin to what has sometimes been called pragmatic/discourse/conversational inferencing, but the analyst should take care not to over-interpret the data or bring his/her preconceived analysis to the data. This is very hard to do, but is crucial for rigorous, empirical work with authentic language use since on the one hand use is very rich and complex and on the other hand we want our analyses to come out of the data, not to be imposed on them.

Because of this richness and complexity, discourse analysis is a multi-/inter-disciplinary field made up of traditions from many different disciplines that often don't overlap, for example, linguistics, cognitive psychology, anthropology, and sociology, to name only a few (for overviews of discourse analysis, see Coulthard 1977/1985; van Dijk 1997a, b; Schiffrin 1994; Georgakopoulou & Goutsos 1997; Jaworski & Coupland 1999; Johnstone 2002; Schiffrin, Tannen, & Hamilton 2001; McGroarty 2002; Phillips & Hardy 2002). Here, we will focus on those empirical approaches that, in combination and incorporating both qualitative and quantitative methods (see Lazarton 2002), have proven to be best for our own work – best in terms of what we've studied, what questions we've asked, and also where the data led us (see Wood & Kroger 2000 for a different combination than ours).

2. Conversation and pronouns: Linguistic approaches to discourse

We'll start with Fonseca-Greber and Waugh's work on spoken French, which we began separately because we were interested in the properties of spoken French and knew that it is very different from written French and indeed from what is taught in French classes here in the U.S., and then decided to do joint work given our common interests. We decided to focus on the subject pronouns because we knew from our own experience that there were interesting differences between spoken and written French in this regard. Indeed, previous published work by others had already suggested that the subject pronouns of French that are presented in reference and teaching grammars and used in the written language do not reflect the everyday speech of educated, middle-class native speakers (for an overview of the French situation, see Fonseca-Greber 2000; Fonseca-Greber & Waugh 2003a, 2003b; for accessible work on English pronouns see Mühlhäusler & Harré 1990; Wales 1996). Native speakers of French, like speakers of any language, are unaware of how much the spoken language is different from the written language and are not conscious of their own linguistic usage, especially in the informal spoken language. Indeed, as is well known, much of language structure and language use is unconscious or at the very least subconscious, so it was clear to us that asking native speakers what pronouns they use, when they use them, what they mean, etc. would not work. And there was no really rigorous study of this situation based on a corpus of authentic use in ordinary conversation. We therefore set out to address this issue, with the aid of corpora of authentic language use and the general findings of corpus linguistics (although most corpus linguistics is focused on English – there is little work, even now, on French). We also were inspired by anthropological and (socio)linguistic discourse analysis, which taught us that work on any part of language – such as pronouns – needs to take into account the larger discourse of which they are a part and the many communicative functions they may fulfill. Pronouns are deictic (shifters, also called indexicality) – e.g., in their meanings, they make reference to their use in speech event contexts (e.g., *I* means the person who is uttering *I*) – and so we were also influenced by work on deixis and indexicality in linguistics. We read as well research in the fairly new field of pragmatics, especially linguistic/ functional/ discourse/socio-pragmatics – and, in particular, those approaches that study language use in relation to the discourse/textual/ social/speech event context in which it is used (see Lyons 1977; Levinson 1983; Anderson & Keenan 1985; Givón 1989; Grundy 1995; Yule 1996; Verschueren 1999). Because we suspected that we would find changes in progress for the pronouns (the written language reflecting the older stages of the language, the spoken language the newer stages), we therefore took into account the two fields that look seriously at language change: 1) historical linguistics, and especially the area of grammaticalization, that studies how grammatical categories (like pronouns) evolve in meaning and use (see Heine, Claudi, & Hunnemeyer 1991; Hopper & Traugott 1993; Bybee, Perkins, & Pagliuca 1994; Bybee 2003); and 2) sociolinguistics and sociology of language, that focuses on different varieties of language use in a given speech community and on language use, ideology and social categorization, was also relevant. This was also coupled with work on ideology and identity that addresses the social and the cognitive/psychological forces that influence both (see Downes 1984/1998; Fasold 1984, 1990; Chambers 1995; Holmes 1992/2001; Hudson 1980/1996; Labov 1972, 2001; Coupland, Sarangi, & Candlin 2001).

In order to study pronouns (and other grammatical issues), we decided to use spontaneous, naturally-occurring face-to-face everyday conversations, audio-taped under naturalistic conditions and transcribed by the authors and/or others under their guidance. There are two corpora represented here: Waugh's French corpus (±77,000 words) of Everyday Conversational Metropolitan French (ECMF) and Fonseca-Greber's Swiss (French) corpus (±117,000 words) of Everyday Conversational Swiss French (ECSF), both of which were created for the purposes of this kind of investigation. In some cases, findings from either one of these corpora are reported here; in other cases, the two corpora were combined into a single larger corpus of Everyday Conversational European French (ECEF), containing ±194,000 words, spanning 15 conversations and including the speech of 27 educated middle-class speakers: 11 men, 16 women, 10 over 40 years old and 17 under 40. This larger corpus allowed more robust (significant) quantitative results. Since care must be taken before using only one corpus or pooling corpora if the results are claimed to be valid more generally, we noted that there are many French (socio)linguists who consider the French of France, Belgium, and Switzerland to have no significant differences in grammar. Moreover, we did an empirical examination of our two corpora for the features we were interested in analyzing and found that there was little/no difference between them – and in cases where there was a difference, we did not pool the results.

The inventory of pronouns for the spoken language with their meanings was established using traditional, discourse, and corpus linguistic techniques of searching for the forms and establishing their meanings through a detailed examination of their use in context. Table 1 gives the written language pronouns that are found in any reference grammar/language textbook; Table 2 gives the spoken language pronouns as we established them through our work.

Now, there are many differences between the two tables and we had two major decisions to make: 1) what to focus on for our research, and 2) what to report on in this chapter, which, given its brevity, could only capture part of our research. We decided to start with a particularly striking quantitative empirical result that had sparked our (and others') interest in subject pronouns at the very beginning: namely the status of the sub-

Table 1. Subject personal pronouns of written French

Form of the verb	Pronouns (Singular)	Pronouns (Plural)
1Sg.	je, j' 'I'	1Pl. **nous** 'we'
2Sg.	**tu*** (familiar) 'you'	2Pl. **vous** (familiar Pl., formal Sg. and Pl.) 'you'
3Sg.	il (masc.), elle (fem.) 'he, she, it'	3Pl. **ils** (masc.), elles (fem.) 'they'
3Sg.	**on** 'one' (indefinite)	

* The forms in bold are discussed in this chapter.

Table 2. Subject personal pronouns/prefixes of spoken French

Form of the verb	Pronouns (Singular)	Pronouns (Plural)
1Sg.	je, j' 'I'	1Pl. [**nous** 'we' rarely used]
2Sg.	**tu**, t'* (familiar Sg„ Pl) 'you'	2Pl. **vous** (formal Sg., Pl.) 'you'
2Sg.	tu, t' indefinite 'you'	
3Sg.	il (masc.), elle (fem.) 'he, she, it'	3Pl. **ils** (masc., fem.), elles (fem. Pl) 'they'
3Sg.	**on**- 'we'	
3Sg.	**on**- 'one' (indefinite)	3Pl. **ils**- (masc.) [i, iz] indefinite 'they'

* The forms in bold are different in the spoken language and the written language as discussed in this chapter.

Table 3. Loss of *nous* and replacement by *on* in ECEF

n. = 1348	Tokens	Percentage
Nous 'we'	13	1%
On 'we'	1335	99%

ject pronoun *nous,* which is 'we' in the written language, but the spoken results show that the meaning 'we' is now given by *on*, not *nous*, which has all but disappeared from the spoken language as a subject pronoun (Fonseca-Greber 2000; Fonseca-Greber & Waugh 2003a, 2003b).

Example (1) shows the use of *on* to mean 'we'.

(1) **on**-s-est mariés deux fois ouais…ici et aux Etats-U – ouais (S)
'**we.**got.married twice yeah…here and in the United St – yeah'
[note: according to our empirical results, all the subject pronouns studied here are actually prefixes and will be transcribed with a hyphen to show this status and glossed as part of the word in the English translation. This issue was explored in some depth in Fonseca-Greber, 2000, and reported on in less detail in Fonseca-Greber and Waugh, 2003.
'United St –' indicates that in the French original, *Etats-U* – is an incomplete word]

99% replacement of *nous* by *on* is an overwhelming, highly significant empirical result – and a surprising one, since when native speakers talk about the use of *on* for 'we', they claim that it is a minor phenomenon and attribute it to uneducated, lower class speech (but, note that our corpus contains the speech of educated, middle-class speakers). Moreover, the extent of the use of *on* has also gone unnoticed by many linguists: this is no doubt because the use of *nous* for 'we' is fully acceptable and grammatical in written French, and *nous* is still used in conjunction with *on* for emphasis or contrast. Thus, when asked for grammaticality or acceptability judgments of fabricated utterances by linguists who rely on such traditional, non-empirical approaches, literate native speakers, who are aware of the conventions of written French as well as the prescriptivist norms for the spoken language, accept these forms. For these native speakers, 'acceptable' or 'grammatical' could be interpreted as 'good' (what they learned at school), not as what they actually say. Additionally, when such informants respond with *nous* for 'we', perhaps they are giving us not the *nous* of written French *nous parlons* 'we speak' but rather the former independent – but current optional, emphatic or contrastive – subject pronoun *nous* of *nous on-a* 'we we-have', since if *on-* functions as a prefix, it may not be cognitively accessible to speakers independently of the verb stem *-a*. Whichever of these explanations may more accurately reflect the cognitive processes speakers engage in when responding to a grammaticality judgment task, a corpus of actual language use allows researchers to uncover previously undetected, even major patterns of language use in a way that intuitive judgments about the language cannot. But we, as researchers must be sensitive to our own interpretive assumptions.

This brings up the issue of the small number of tokens of *nous* listed above (13 tokens, or 1%): in fact, all of these tokens occur in the French corpus (ECMF), not the Swiss one. Through an analysis of the data and correlation of the data with what we know about the participants in our corpus (gained by a close reading of the transcripts of all the conversations, as well as participant data-forms), we found that the majority of the 13 tokens of

Table 4. Relative Frequency of Use of *on*- for 'we' and 'one' in ECEF

n. = 1749	Tokens	Percentage
On = Indefinite 'one'	100	5.7%
On = Vague*	314	18%
On = Personal 'we'	1335	76.3%

* Vague=could be interpreted as either Indefinite or Personal in the context.

nous are used by educators (cf. the French Academy, with its heavy emphasis on 'correct' usage in schooling) and thus tend to be related to linguistic conservatism. However, this is true for France only: some of the Swiss speakers are also (retired) schoolteachers but they don't use *nous* (there is no French Academy in Switzerland, and the Swiss in general may be less worried about 'correctness'). In addition, some of the participants in ECMF are of North African background and their *nous* usage may reflect their desire to show their knowledge of formal/written/high prestige norms and to assimilate in French culture. So, even if they are native or all-but-native speakers of French, they may make less use of the range of registers that are available and in particular may cling to the more formal uses of the language.

These findings about *on* bring up a number of questions, since in the written language *on* is **the** way of giving indefinite meaning and the widespread intuition is that the situation is the same in spoken French. Indefinite pronouns are typically used when the speaker does not want to be or cannot be definite about the reference of the pronoun. In English, for example, some indefinite pronouns are 'one' in the formal language ('one shouldn't do that'), and, in informal usage, 'you' ('to go from Tucson to Phoenix, you take I-10'), 'they' ('they say it's going to rain tomorrow'), and 'we' ('we're going to make a mess of the planet if we don't start conserving energy now'). This led us to investigate a new set of research questions that we had not started out to look at but became vital as soon as we saw the quantitative results: how often is *on* used as an indefinite?; are other personal pronouns used as indefinites?; if so, what is the difference between them conceptually, pragmatically, discursively?; what would lead the speaker to use the one rather than the other?

Our first task was to establish a base-line, that is, how often *on* is used for indefinite meaning, and the answer, based on ECEF, was that it is used in only a small number of cases for indefinite meaning, as displayed in Table 4.

Now, work in linguistic semantics shows that most grammatical and lexical items in a language have multiple meanings (polysemy) as they are used in different contexts and situations. But, in grammar, especially, there is typically one meaning that is the basic (core, nuclear, prototypical) meaning (see Jakobson 1936/1982; Comrie 1976: 11; Taylor 1989/1995; Ungerer & Schmid 1996; Lee 2001), which is assumed to be contextually the least conditioned and cognitively the most salient, and is quantitatively calculated to be present in about 2/3rd (or more) of all tokens. According to our empirical findings, Table 4 shows that the basic meaning of *on* is not the indefinite (5.7%) but its new personal meaning of 'we' (76.3%), which is a complete reversal from its earlier meaning and from its meaning in the written language. Example (2) shows one type of indefinite meaning of *on*:

(2)	parce moi	**on**-m-a fait	légaliser	le	mariage (S)
	'cuz for.me	**they**.had.me	authenticate	the	wedding'

[note that the object pronoun *me* (here *m'* because it precedes a vowel) is, like the subject pronoun, a prefix. All object pronouns are transcribed as prefixes in French and glossed as part of the word in English.]

In (2), *on* means some undefined authorities, unknown to 'me'. Note that the presence of 'me' *m'* as the direct object of the verb means that it is impossible to interpret this use of *on* as 'we'. This is typical of non-basic (peripheral, marginal) meanings: they need contextual support for their interpretation and this gives empirical justification for the analysis. In example (1) above, the interpretation of *on* as meaning 'we' is obvious from the context, since the speaker is talking about herself and her husband and their two weddings to accommodate family in the U.S. and in Switzerland. Table 4 contains a third category – vague *on* – which accounts for almost a fifth of the tokens; example (3) is one case from our data.

(3)	C'est	ridicule	**on**-a même pas	besoin de permis (K)
	'it's	ridiculous	**one.**doesn't/**we.**don't need	to have a permit

[to buy a gun in the United States]'

The general context here is that K, who lives in the U.S., is talking about violence in the U.S. This utterance could be interpreted to mean either indefinite 'one' (anyone, everyone=any American, all Americans), or as personal 'we' (we Americans) in the context. The discovery of this vague meaning is a contribution to work on both language change and linguistic interaction. On the one hand, it shows how the gradual semantic drift from indefinite to personal, seemingly so implausible at first glance, could have taken place; on the other hand, vagueness about meaning makes *on* very versatile, since it allows the interlocutors not to be specific about what they're referring to and thus offers the broadest range of possible interpretations (see Channell 1994 on vague language). Indeed, speaker and hearer could have different cognitive understandings of the intended meaning, but even so, it is obvious from a close analysis of the data that the flow of conversation is not affected by this vagueness at all.

The empirical results of Tables 3 and 4 raise many new questions, and we decided to continue our quest about indefinite meaning, since, if *on* is only marginally used now in its original indefinite meaning, it seems as if it probably is no longer the preferred way of expressing indefinite meaning, but this has to be proven empirically. The results in Table 5 show that *on* accounts for barely over a quarter of the indefinite (including vague) tokens (or only 8.5% if we leave out the vague tokens). So, how do you give indefinite meaning in French?

Table 5. Indefinite Meaning: 'one', 'you', 'they' in ECMF

n. = 1489	**Tokens**	**Percentage**
On (Indef. 'one' & Vague)	414	27.8%
Tu (Indef. 'you')	918	61.7%
Vous (Indef. 'you')	18	1.2%
Ils (Indef. 'they')	139	9.3%

By reading the literature in French linguistics, we knew that there were others who had already said that the French equivalents of 'you' (*tu* and *vous*) were used as indefinites, so we decided to focus on these two pronouns and their uses in our corpus. What we found was that the vastly preferred indefinite is *tu* 'you [familiar/solidary]', accounting for over 60% of the indefinite tokens (or, 78%, if we take out the vague tokens of *on*); *vous* 'you [formal/plural]' is rarely used in this meaning in our corpus, and so we didn't focus on it in this study. One type of indefinite usage is exemplified by (4), where *tu* means 'one', 'anyone':

(4)	Même	dans	les villes	comme	Cincinnati	ou	j-sais pas	des villes	qui	sont	un	peu
	'Even	in	cities	like	Cincinnati	or	I.dunno	cities	that	are	a	little

	plus	perdues	que	Washington	**tu**-trouveras	beaucoup	plus
	more	isolated	than	Washington	**you**'ll.find [one.finds]	many	more

de gens obèses. (K)
obese people'

It seems to us that this preference for indefinite *tu* could be related to the rise documented by (corpus and socio-) linguists in the use of *tu* (vs. *vous*) as the personal form of address among the young across both sexes, all social classes, educational backgrounds, etc. (see the classic text by Brown & Gilman 1960 on this issue – although we should also say that a study based on corpora with many more uses of the *vous* pronoun than we have in ours might prove interesting). We also think that this rise in the use of *tu* could be related to the spread in France and in Europe more generally of an ideology of egalitarianism, less distance, more solidarity, and less formality between interlocutors. This spread has been documented as occurring since the change to democratic governments in Europe (over a few centuries), but may have received new impetus from the general social changes that have taken place since the 1960's. All of these speculations cannot be addressed by our corpus and await further empirical work, but they promise to be fruitful ground for the strong ecological tie between conceptual/ cognitive and social phenomena.

In addition, the spread of *tu* for indefinite meaning is also not just a French phenomenon, since empirical work on other European languages in corpus and sociolinguistics (especially work on the spread of linguistic features across languages in contact with each other – see Thomason 2001) shows that the equivalents of 'you' in these languages are also increasingly used as an indefinite. Now, since English, the globally dominant language with high prestige, also uses 'you' as an indefinite, there is a possible ideological influence on usage in a variety of languages in contact with English and with each other. Here, there are many avenues of research that should be explored further.

However, we decided to continue work with our corpora and, in particular, we wanted to look further at a third pronoun used for indefinite meaning, namely, *ils* 'they', with 9.3% of the indefinite tokens (or 11.8%, if the vague tokens of *on-* are excluded). Here too we noted a parallelism with indefinite 'they' in English (although we don't have the same empirical evidence for direct ideological influence from English as we do for 'you'). In Table 6, we summarize our empirical findings for *ils* in both its personal and indefinite meanings. Example (5) is one type of indefinite *ils*:

Table 6. Frequency of Personal vs. Indefinite Uses of ***ils***

n. = 341	Tokens	Percentage
Ils = Personal 'they'	198	58.1%
Ils = Indefinite 'they'	139	40.8%

(5) **Ils**-te-donnent une bourse de transport de l'equivalent de 30 dinards. (K)
'They.give.you a transportation scholarship the equivalent of 30 dinards.'

We see, then, that there are three different indefinites in spoken French: most frequent/newer *tu*, less frequent/older *on*, and less frequent/newer *ils*. At this juncture, we made the decision to explore these three further in terms of their meanings/usage and so we did a fine-grained semantic and pragmatic analysis of their uses, and came to the following conclusions:

(a) *Indefinite* ***on:*** more formal than **tu** or **ils**; general and wide-ranging in its use; it is the unmarked of the three (Waugh 1982); it can express solidarity or distance or be neutral (see Brown & Gilman 1960); it can be positive or negative or neutral; potentially lighter, face-saving because of its possible neutrality (where face is defined as "a socially attributed aspect of self that is temporarily" construed in verbal interaction with others" (Watts 2003: 125), as first explored in depth in the classic text by Brown and Levinson (1987)). Because *on*- is used in its basic meaning for 'we', it has a tendency to include 'we' in some (but not all) contexts. In example (2) above, indefinite *on*- is exclusive, distancing, possibly negative, and as noted above, does not include any nuance of 'we' due to the presence of the object prefix *m*- 'me' in the same sentence, which rules out 'we' as a possible meaning; in example (3) above, *on* could be either inclusive or exclusive.

(b) *Indefinite* ***tu:*** less formal; inclusive=potentially inclusive of the interlocutors as virtual/potential participants in the verbal process; highly personalized because it can include and express solidarity between speaker and addressee; strives for communicative involvement of the addressee in the topic of the utterance; usually shows positive face and often gives a positive evaluation of the verbal process. In example (4) above, *tu* functions as an inclusive indefinite: even though only K has been to America, by using indefinite *tu* he includes his interlocutors as virtual participants in the experience, in an act of communicative solidarity. He thereby possibly achieves their greater communicative (including cognitive and emotional) involvement in what he is talking about. The same is true of the use of *tu* as the object in (5).

(c) *Indefinite* ***ils:*** less formal; exclusive=potentially exclusive of the interlocutors as virtual/potential participants in the verbal process; expresses distance between the speaker and addressee on the one hand and the topic of the utterance on the other hand; often gives a negative evaluation of the verbal process. In example (5) above, K is referring to unknown authorities that don't include speaker and addressee. Here the distancing and possible negative evaluation is counteracted by the presence of the object pronoun *te* 'you', and the topic (being given money to study Tunisian Arabic in Tunisia by the Tunisian authorities).

In example (6),

(6)	à	Genève	le résiné	**ils**-le-font: (S)
	in	Geneva	fruit.juice.pie	**they.**make.it'

indefinite *ils* metonymically means 'those people who live in Geneva', from which, given the distancing of *ils*, the usual pragmatic inference would be that the speaker doesn't live in Geneva. However, given the ethnographic information about the participants and participant-observation, we know that the speaker **does** live in Geneva. Now, as we saw in example (4) earlier, establishing solidarity in conversational interaction is an effective rapport-building strategy. We therefore interpret this usage in this way: by framing the Genevans as the 'other' ('those people in Geneva'), her interlocutors who are from the adjoining canton of Vaud can interpret her utterance as aligning herself **with** them and not against them.

In the following example, the speaker uses the personal *tu* in its new plural meaning several times (and the discourse marker *t'sais* 'y'know' – discourse markers are expressions like English *well*, *but*, *oh*, *y'know* that have discourse, social, cognitive and expressive functions, but don't actually refer to anything concrete in the external world – see Schiffrin 1987, 2001). She uses as well the indefinite *ils* (and its object counterpart *leur* 'them'). In this interaction with several interlocutors including an American who is married to a Swiss, the speaker shows solidarity/commonality (positive face) by the use of *tu* with everyone, including the American, and the indefinite *ils* when talking about Americans in general: in this way, the American is being constructed as Swiss, not one of those finicky Americans who don't like trying new foods:

(7)	**t**-as déjà remarqué	qu'**ils**-sont pas	très	courageux
	'Have **y'**[plural]noticed	that **they** [Americans] aren't	very about	brave

au point de vue	bouffe	non plus?	... t'sais	quand	**t**-essaies...quand
about	food	either?...	y'know	when ...	when...**y'**try...when

tu-leur-prépares	quelque chose	d'européen. (S)
y'prepare.for.them	something	European'

In contrast with the virtual inclusiveness that indefinite *tu* creates, indefinite *ils* tends to exclude speaker and hearer and can be used to create distance from those (indefinite people) being talked about.

At this juncture, there were many empirical paths we could have taken – for example, there is much more to be done about the differences between *on*, *tu* and *ils* as used by different speakers, in different context, about different topics, etc. Indeed, there is a whole literature on indefiniteness in English that could be used as a comparison to French. However, we decided to focus next on a research question that arose by working closely with the data and that was due to the special nature of some of the participants in the French data – namely, the fact that some of them were North African in ethnicity and thus they presented very interesting issues for language use, identity, ideology, etc. In particular, we found that the *on*, *tu,* and *ils* indefinites can be used in conjunction with each other to create or shift identities (for the question of identity, see work in interactional sociolinguistics cited above and in intercultural communication (Scollon & Scollon 1995/2001)) or to align oneself with or shift ideologies throughout a conversation (ideology is explored in intercultural communication, linguistic anthropology (Duranti 1997), sociolinguistics, and critical discourse analysis (CDA) (Fairclough 1995; van Dijk 1998, 2001a, b; Fairclough & Wodak 1997; Wodak & Meyer 2001). We decided to concentrate on one especially interesting conversation tape-recorded in France in 1998 between Karim and three others (Karim is a pseudonym – it is standard to use pseudonyms in this type of work in order to protect participants' confidentiality) that led us to further exploration of this issue. We should say that when we began our research we had no intention of looking at indefinites at all – that arose because of our findings about *on* – and we certainly had not envisaged that we would make a foray into the question of identity/ideology – but, as is often the case with empirical work, the data were so compelling that we felt drawn to consider these issues.

As revealed in the ethnographic data as well as in this conversation and others in which Karim participates, he regularly constructs himself as having three ethnic/national identities. He is a French citizen; his grandparents live in France, and his parents used to live in France and still have close friends there whom they visit on a regular basis. At the time of the recordings, he's visiting friends and relatives in France during the summer (sometimes with, sometimes without his mother) – in the case of this conversation, he is with friends, and without his mother. At one point in this conversation, he says "I'm from a French family". But he also makes it clear that he also considers himself to be an American; his parents have lived in the U.S. for some time, near Washington, D.C., he did some of his secondary education there, and he is an undergraduate in an elite liberal arts college in the Northeastern U.S. (he's from an upper middle class family). He gives his interlocutors lots of information about all of this. And, he is of Tunisian descent, and at one point he says "I'm Tunisian-American". In addition to his three ethnic identities, Karim's linguistic situation is also complex: he is a fluent bilingual speaker of both French and English (the other three in the conversation are native speakers of French and don't speak English), and just before this conversation he had gone to Tunisia for a month to study Tunisian Arabic (which he doesn't speak). As we will see, he shows his shifting identities and his ideology(ies) through his use of the indefinite pronouns.

Early in the conversation, Karim talks about his language training in Tunisia, and part of what he says was given in example (5) above, incorporated in example (8) here:

(8) Karim: **Ils**-**te**-paient la moitié de **ton** voyage **Ils**-**te**-paient les cours.
'**They**.pay.**for.you** half of **your** trip. **They**.pay.**for.you** course fees.
Ils-**te**-donnent une bourse de transport de l'équivalent
They.give.**you** a a transporation scholarship of about
de 30 dinards. [...] **On**- **on**-fait pas mal avec. Et euh **on**-**te**-donne
30 dinards. [...] **One. one.**can get along on that. And uh **they.**give.**you**
une chambre dans un foyer universitaire.
a room in a dormitory.

By the use of *tu* ('you' and the object and possessive adjective equivalents *te* 'you' and *ton* 'your' respectively), Karim indexes himself as Tunisian and at the same time creates solidarity with and involvement of his interlocutors (none of whom is Tunisian). *Ils* 'they' in (8) means the anonymous, faceless Tunisian authorities. He differentiates them from the university people who gave him a room in the dormitory, since he uses *on* in the last sentence when talking about them and the inference is that he met them (had face-to-face interaction with them). The next to last *on* (in *on- on-fait pas mal avec* 'one. one.can get along with that' – the repetition of *on* is known as a 'false start'), rendered awkwardly as 'one' in English, is vague: it could mean 'we' (the participants in the course, or even more generally Franco-Tunisians living abroad who go there), or it could mean 'anyone visiting Tunisia'. But, as said earlier, it doesn't matter what it means exactly, since the conversation continues without any problems and it may be that Karim himself couldn't say which one he might have meant – perhaps he meant them all.

Later in the conversation, Karim constructs himself as being an expert on America, and when he talks about violence in America and how easy it is to buy a gun, he says (part of this was example (3) above):

(9) K: C'est c'est ridicule. **On**-a même pas besoin de permis t'vois c'est juste
'It's it's ridiculous. **One.**doesn't even need a permit y'see that's right
< – > **on**-achète son arme et **on**-sort C'est ridicule. Et après **ils**-sont
< – > **one**.buys the gun and **one**.leaves It's ridiculous. And then **they**'re
étonnés d'avoir des trucs, des tueries comme ça.
surprised to have stuff, killings like that one."

Here, the use of vague/neutral *on* indexes the fact that buying a gun in America is an everyday occurrence that anyone can do (and this is in contrast with France, and his interlocutors know that the American situation is completely different from the French one). Karim shows his distance from the American situation – and from America and Americans – by stating twice "it's ridiculous" and by using the indefinite *ils* in his last sentence, referring to (those faceless, anonymous) Americans from whom he distances himself ideologically: 'not me', he implies; 'in this respect, I'm not an American'.

In a final example, Karim is talking about obesity in the U.S.:

(10) Alors **ils**-font attention alors en ville. Quand **tu**-vois les obèses
'So **they**.pay more attention [to weight] in the city. When **you.**see obese people
on-peut assez facilement réparer s'**ils**-sont de la ville ou pas...
one.can pretty easily figure.out if they're from the city [Washington] or not. . . .
Même dans les villes comme Cincinnati ou j'sais pas des villes qui sont un peu
Even in cities like Cincinnati or I don't know cities that are a little
plus perdues que Washington **tu**-trouveras beaucoup plus de gens obèses.
more isolated than Washington **you**'ll.find many more obese people.
Et ça c'est clair.
And that's for sure.

Here, again, Karim reveals his shifting identity and his ideology through his use of the indefinite pronouns. When he uses *ils* in the first sentence, he means 'those Americans' ('not me'), even though elsewhere in this conversation he very much identifies himself as an American; in this same sentence he makes reference to 'the city' (meaning Washington, D.C.), which he has talked about earlier at length as being 'my city' (because he grew up there, his parents live there, and he goes there often). In the next sentence (and in the next to last sentence), he uses *tu* as a conversational involvement strategy with his interlocutors, who could imagine themselves there. But then his use of the more neutral *on*, which could be rendered better in English through the impersonal – 'it's pretty easy to figure out if they're from the city or not' – suggests that anyone could tell whether people are from the city or from somewhere else (e.g., American tourists visiting Washington) by how obese they are.

In these two excerpts about America, we can see that when it comes to some of the stereotypical characteristics of Americans that the French (like his interlocutors) don't like about Americans (buying guns and being obese), Karim makes it clear that while he lives in America and in certain ways he is an American, he is ideologically not American – he's not one of them.

We started this section with the issue of what the pronouns of spoken French are and we have seen that by using a variety of different qualitative and quantitative empirical methods inspired by many different traditions, most of them related to linguistics in some way, we have not only been able to make discoveries about forms, meanings and uses of the pronouns but also to address a variety of social and cultural issues important for cognitive analysis, such as establishment of solidarity with or distance from the addressee by the speaker, positive and negative face, indexing of linguistic and cultural identity, and showing one's ideology – all through a fine-grained analysis of the use of indefinite pronouns in conversational interaction. Now, given our data, there are many more empirical issues we could have raised – and indeed we have addressed some of them in other contexts (see Fonseca-Greber 2000; Fonseca-Greber & Waugh 2003a, b), but there is much more that needs to be done. A corpus of authentic data is like a treasure-trove and working with one can take years. However, if we wanted to address questions having to do with the use of specific lexical items (one of the traditional ways in which larger corpora are used), we would have to gather more data. We should also note that Waugh's corpus contains some spoken academic discourse (university lectures) and what are called 'service encounters' (e.g., talk between a cheese store owner and customers, between a pharmacist and customers) – and these too can be used for a variety of different empirical questions, which have yet to be explored.

3. Discourse and interactional accommodation: CCCA

The discussion in this section concentrates on my own (Vickers') English corpus (±150,000 words), focusing on conversational interactions between native speakers (NS) and nonnative speakers (NNS) of English in the U.S. academic community (see Vickers 2004). My study demonstrates ways in which NSs and NNSs interactionally accommodate to each other as they try to perform a task together, and, in particular, instances in which interactants successfully come to work within a shared interpretive frame, i.e., ways of knowing and of experiencing the world (Gumperz 1982). Furthermore, the discussion will demonstrate that coming to work within a shared frame has implications in terms of the power relations between NSs and NNSs in this speech community. I will discuss in this section the research process, including the challenges of collecting a corpus for the purpose of studying a specific set of research questions, as well as the final product of my study.

To begin, I will discuss the challenges involved in finding the appropriate context in which to collect data related to a specific set of research questions. It is not always easy to find a context friendly to the everyday presence of the researcher, and it is crucial that the researcher be present on a regular basis when collecting a culturally contextualized corpus of this sort. In the case of my study, I originally wanted to do research in the context of a college of business in a major American university. I had hoped to investigate Chinese and Korean students' interactions with American students during teamwork in this context. Unfortunately, though the college of business was open to my conducting experimental research in which students had contact with me in a laboratory type situation, they were not comfortable with me naturalistically observing their teams. Therefore, I had to find a different context in which to address my research questions. The department of Electrical and Computer Engineering (ECE) as well as Family and Consumer Sciences enthusiastically welcomed my research. The ECE department offered me access to a large undergraduate course, while Family and Consumer Sciences could only offer me access to a small graduate seminar.

After considering these options, I decided to work with the ECE department and thus the corpus I'll be discussing was gathered in the context of a large undergraduate design course, in which typically 70% are NSs and 30% are NNSs. Though there were quite a few NNSs in the course, I was unable to recruit multiple Chinese and Korean participants. As a result, rather than focusing on Chinese and Korean students' interactions with American students, I was forced to revise my research questions to encompass a study that would examine NS-NNS interaction in the ECE speech community, regardless of country of origin. Therefore, my research questions had to be revised to reflect the population that I was actually studying, one in which the NNSs came from a variety of cultural backgrounds. The actual data that I had access to led to the revision of my research questions.

In the ECE design course, all students are required to work together throughout the academic year on teams, formed by the professors according to the students' technical interests and background, to design operable electronic devices. The situation is high stakes, since students' graduation is dependent on the design of a successful device. As such, the ECE team meetings represent a naturalistic context in which to study NS-NNS teamwork. The focus of this study was shaped by the context of the study, the data, and thus the salient issues in this high stakes, competitive, and quite male context.

To collect data, I employed microethnographic methods, as developed in the ethnography of communication. Besides video and audio taping of the team meetings, I engaged in participant-observation: I observed the classes, I was at the team meetings and I was present at the ECE portfolio evaluations. I also conducted participant playback sessions (where I showed the video tapes to the participants and asked them to comment on certain factors of the interaction). The objective was to gain an understanding of the speech community into which the ECE students were being initiated and of the ECE students' participation within that speech community, so that I could understand what was happening in the team meetings, which themselves constituted communities of practice. Observation of the course lectures allowed me a certain amount of insight into the forms of talk (Goffman 1981) in the ECE speech community – these forms of talk are quite specialized, including extensive use of technical jargon. Observation was also informative in terms of the course objectives and the standards of professionalism in communication that were expected in the ECE team meetings.

One important consideration when engaging in ethnographic work, particularly in an unfamiliar speech community, as in the case of my study in the ECE community, is to allow the salient interaction types and implications of those interaction types in that particular community to emerge from the data. Especially when much is unfamiliar as the research begins, initial research question must be a) quite broad and b) flexible. I had to be willing to revise the research questions based on the data. The data drove the research questions, and revision of research questions was a recursive process that happened in combination with data collection and analysis.

As I collected and began to analyze the data, it became clear that there were particular interaction types that allowed teammates to accommodate to each other so that they could engage in negotiation. It also became clear from the data that NSs and NNSs had different ways of accommodating. Thus, one of my research questions became, "What strategies do NSs and NNSs employ to accommodate to each other?" Another became "What implications do these strategies have for the hierarchical structure of the team?" I formulated these research questions based on what the data was showing me. As I engaged in data collection and analysis in investigation of these questions, it was evident that I would need to involve multiple theoretical approaches to fully analyze the data and address my research questions, as I will demonstrate below.

In conducting data analysis, at the micro level, consisting of fine-grained analysis of participants engaged in face-to-face interaction, I incorporated analytic methods from CA, interactional socioliguistics, and corpus linguistics. At the macro level – analysis of the larger context within which the participants operate – the work was inspired by ethnography of communication and in particular CCCA (Moerman (1988) discussed above). The collection of multiple instances of the same speech event in the same cultural context within and across teams allowed for the kind of systematization conducive to both quantitative and qualitative corpus analysis.

On the level of micro interaction, the most important influence from CA is the necessity to examine the interdependency between utterances in order to understand how utterances are contextualized within a conversation and how conversants' contributions are constructed within the conversation at hand. By examining utterances in relation to surrounding utterances in the context of an ongoing conversation, I was able to locate conversational sequences, in this case, accommodation sequences. The accommodation sequences explored in this analysis contain in the first utterance either the solicitation (explicit request) for a clarification by a participant in the speech event or an explanation by a participant in the speech event that was not solicited by the other participants. The clarifications and explanations are of four types: an assumption, an opinion, technical content, and linguistic content. I made decisions about coding these sequences in light of the conversational context in which the speaker made the utterance and according to the way the utterance was understood (uptaken) by the interlocutor. In other words, while in some cases the intention of the speaker may be different from the understanding (uptake) of the addressee, the flow of the discourse is determined more by how the addressee understands what the speaker has said and this in turn determines how the addressee responds to what the speaker has said – and thus, the focus here is on the co-construction of what the discourse is about, since that is what the data show (for a discussion of co-construction, see Jacoby & Ochs 1995). In other words, from a cognitive perspective, what is important is how the interactants understand what was said and how it should be understood, not just what the speaker might have meant or intended.

Examples 11 and 12 are solicited clarifications; they are solicited clarifications rather than unsolicited explanations because in (11) P and in (12) R ask for clarification through their direct questions – italicized in the example. (Other clarification sequences not exemplified here are Solicited Clarification of Opinion=SCO, and Solicited Clarification of Linguistic Content=SLC).

(11) *Solicited Clarification of Assumption Sequence* (*SCA*)

G: I don't know if it's worth it maybe he'll give me some of the old parts and maybe that'll cut it for what we need and it'll at least give us something to test in the lab

P: *well what does that mean for like part one b does that mean if you can't get the parts until January that we're putting it off until January*

G: yeah that's what that means

(12) *Solicited Clarification of Technical Content Sequence* (*SCTC*)

G: I mean it wouldn't be that bad if we could run this off five volts I don't know

R: what what's the matter with five volts

G: [it's]

P: [it's] not a good battery voltage

G: yeah you got a 1.5 volt battery so you need four of 'em

R: yeah that's true

G: three volts you only need two double A

[Notice that the brackets around G's [it's] and P's [it's] means that they overlap with each other=talk at the same time]

Example 12, involving two NSs, George (G) and Peter (P), and one NNS, Ramelan (R) (all pseudonyms), shows the necessity of looking at the interdependency between utterances to identify an accommodation sequence. In their response to Ramelan's initiation of a clarification sequence, George and Peter show that they have uptaken it as SCTC (not SCA): Peter provides the first part of the technical content, and George the second part and thus they are both able to display their technical expertise. Ramelan accepts these responses. This demonstrates that interactants do not come to a particular interaction with all aspects of the interpretive frame predefined (as would be assumed in a pure interactional sociolinguistic analysis) but that the interpretive frame itself is co-constructed in face-to-face interaction, as CA claims, and that these sequences constitute strategies that interactants employ to formulate a shared interpretive frame. This is also the basis for some who claim that cognition is social in nature (see Edwards 1997; Lantolf 2000; Virtanen 2004, among others). Examples 13 and 14 are unsolicited explanations:

(13) *Unsolicited Explanation of Assumption Sequence* (*UEA*)

B: I'm talking about the design you're talking about the design so it's like we are [repeating ourselves]

A: [yeah therefore] we understand ourselves but when you said we should do this I thought with this like I'm specializing hardware I will just this mostly [about the hardware I don't]
B: [no no right so you'll do all]

(14) *Unsolicited Explanation of Technical Content Sequence* (*UETC*)
P: I thought that big one looked like it did but
G: no
P: no it's [surface mount]
G: [surface mount] so what you do is be able to board with a surface mount lay it [on to the pad]
P: oh yeah right

In example (14), George is able to display his own technical knowledge by providing an unsolicited explanation of technical content to Peter. The display of technical knowledge along with an acceptance of that technical knowledge by team members has the effect of shaping the team's interpretive frame. Thus, the micro-level analysis provides strong evidence that displays of technical knowledge are associated with high status on the ECE team, since being able to shape the interpretive frame gives a participant high status.

This brings up the issue of who has high status on these teams. Table 7 demonstrates the percentage occurrence of each type of accommodation sequence, as well as percentage of types of accommodation sequences used by NSs and NNSs respectively.

By examining the frequency of occurrence of accommodation sequences by NS and NNS, I was able to determine major differences in the ways the NSs and NNSs engage in interactional accommodation. In general, NNSs are far more likely to engage in solicited clarifications in general, and in particular they favor SCTC (42% of all their accommodation sequences) and SCA (31%). Native speakers, on the other hand, are more equally divided between the solicited clarifications and unsolicited explanations, and while they do engage in SCTC (20%) and SCA (16%), it is at a lower rate than NNSs and they favor UETC (41% of all their accommodation sequences), while for NNSs the percentage of UETCs is only 8%. This difference becomes important to the construction of the status of NSs and NNSs on the team.

Ethnographic data from the ECE speech community then informs the micro-level analysis because in the ECE speech community, the ability to display technical knowledge is important in achieving high status in the community. Ethnography of communication provides a framework for examining the ECE team meeting speech event as embedded in the ECE speech community. Therefore, coming to know the speech community sheds light on the forms of talk in the ECE team meeting, and how they operate in the speech community. In particular, it is clear that UETC is a powerful, high status move – because

Table 7. Frequency of All Accommodation Sequences

Initiator	SCA	SCO	SCTC	SCL	UEA	UEO	UETC
All	**21%**	**4%**	**27%**	**9%**	**2%**	**6%**	**31%**
NS	16%	6%	20%	7%	2%	8%	41%
NNS	31%	1%	42%	15%	1%	2%	8%

it is unsolicited and because it deals with technical knowledge, which is given high prestige both in the ECE community and in the community of practice of team meetings. On the other hand, SCTC is quite face threatening for speakers, because, by soliciting technical knowledge from their interlocutors, speakers construct themselves as lacking technical competence, and, as a result, as potentially having low(er) status in this speech community/community of practice. Recurring instances in which NSs display (solicited or unsolicited) technical knowledge and NNSs solicit technical explanations then results in the construction of NSs as high status members and NNSs as low status members of the team, not because of their linguistic competence *per se* (although this may be involved) but because of their ability to engage in high status forms of talk within the ECE speech community.

Furthermore, ethnographic data demonstrates as well that it is not simply the NS-NNS distinction that necessarily gives NSs higher status than NNSs, but it is a matter of who has access to high status forms of talk in the ECE speech community in general. In particular, playback session interviews in which corpus participants discuss team interactions while watching videotapes of the interactions shed light on other contextual information, such as the fact that NSs and NNSs have differential access to high status forms of talk within the ECE speech community: NS ECE students typically have the opportunity to gain professional internships which typically last several years and allow them to gain access to high status forms of talk; NNSs do not have this opportunity because of their visa status and thus are less exposed to high status forms of talk.

The integrated analysis of the corpus data shows that people from different cultural backgrounds typically accommodate to each other in the process of face-to-face interaction in the ECE team meeting speech event in order to establish a shared interpretive frame and that there are shared strategies for interactional accommodation that allow participants to construct this frame. However, people from different cultural backgrounds and with different access to high status forms of talk (and experience) tend to employ strategies associated with interactional accommodation with different frequencies and in qualitatively distinct ways. These differences become important to the development of social hierarchy on the ECE team and in the ECE speech community because, in the process of accommodation, much of the construction of team members as possessing technical expertise in particular areas and as being able to make technical expertise relevant to the team's project is accomplished. As such, micro-level interactions construct, co-construct, and re-construct NS team members as high status and NNS team members as low status members of the team and ultimately of the speech community.

4. Interactional patterns of international students: Triangulation

This section is based on my own work (Eröz 2003), which focuses on the interactional patterns of international students with each other and with Americans, in order to determine if international students interact in ways that are different from Americans and if these ways of interacting are a reflection of classroom interactional patterns in their home country. The speech community or communities of practice I analyzed for this project consisted of two mandatory freshman composition courses at a large southwestern American university. The participants were 35 international and 9 American students, some enrolled in an all-international section (IS) consisting of non-native speakers of English (NNS), the others in a mixed section (MS) consisting of both American native speakers of English (NSs) and international students.

In this research study, I focused on the common interactional patterns within four non-American culture groups: (Asian) Indian, Japanese, Chinese, and Middle Eastern. I also investigated international students' perspectives on studying with native speakers in content versus English classes, as well as their sociocultural adjustment difficulties as newcomers in the United States. Furthermore, I analyzed classroom interaction patterns common to three participants from each of the culture groups and compared these to American classroom patterns. My three data collection techniques were ethnography (combined with ethnography of communication), CA, and CCCA. The database for this corpus was compiled through participant (classroom teacher) and non-participant observation as well as video and audio taping of classroom discussion (15.5 hours) and of individual (7.5 hours) and group teacher-student (6 hours) conferences. In addition there were one-on-one interviews with most of the international participants and questionnaires using both Likert-scale and open-ended questions about their experiences in America. I also used my own written data: systematic notes taken during classroom observations in the MS, reflective and retrospective Teacher's Journal entries made after IS class sessions, notes taken during multiple viewings of the individual interviews with participants asking for their reactions, and topic transcriptions of the video-taped data noting the time of the interactions, direction of the interactions (e.g., teacher-to-student), and content of the interactions (e.g., teacher asks if there are any questions). My aim was to do an integrated analysis of multiple data sources – triangulation – in order to come up with relevant, objective, and trustworthy results.

The claim that I addressed through this triangulation was that there are interactional similarities among students from the same culture in terms of how and how much they participate in class, and how they relate to each other and to the teacher in whole class and small group discussion in the classroom, in teacher-student conferences and in interviews. My hypothesis was that these similarities are due to the fact that these students transfer attitudes about the role of the teacher and traditional classroom behavior from their home country to the American classroom. I explored these claims by making comparisons across all the various types of data, searching for commonalities among same-culture groups and differences across cultures.

In order to explain the procedures of triangulation for this data set to address the research questions, I will exemplify the nine stages of my data analysis and interpretation: 1) Question, 2) Database Overview, 3) Assumption, 4) Close Look at Database, 5) Data Source(s) Selection, 6) Example Selection, 7) Interpretation, 8) Literature Comparison, and 9) Conclusion. To illustrate these stages I now focus on the Chinese students in the Asian group and the research question pertinent to investigating similarities among students from the same culture and the cultural influences of this behavior.

1. *Question*: Are there interactional similarities among students from the same culture in terms of how and how much they participate in class and how they relate to the

teacher and each other in communicative activities (e.g., small group discussions and conferences)?

2. *Database Overview*: To address this question, I looked through classroom notes, topic transcriptions of lectures and student-teacher conferences to identify culture-specific patterns of the culture groups represented by three or more participants – in this database, these groups were Asian Indian, Chinese, Japanese, Middle-Eastern, and American. After this, I determined one specific group to focus on: Chinese students.
3. *Assumption*: After the initial overview of data, I observed that Chinese students tended to be quiet during classroom discussions, but collaborative in small group discussions. They seemed modest and respectful towards the teacher; they didn't interrupt or disagree openly with the teacher.
4. *Close Look at Database*: At this stage, I consulted written data sources to note the general classroom interaction behavior of Chinese students; reviewed instances of Chinese students talking in classroom discussions and small group work; and examined their attitudes in individual and group conferences. I revisited the class videotapes when necessary.
5. *Data Source(s) Selection*: I checked and validated assumptions through the use of these following data sources: videotaped classroom sessions; topic transcription of taped sessions and conferences; interview information gathered from or about Chinese students; and questionnaire responses of Chinese students in general.
6. *Example selection*: In order to illustrate the points implicated by the assumptions, I selected striking and clear examples from various data sources:

Data Source 1: Videotaped conferences (examples 15 and 16):

(15) Adam (Chinese NNS in IS): individual conference with the teacher
Teacher (T): You have to give specific examples from the story {writes on his essay}
Adam: A-hm
T: From the story, and then you have to say... you have to explain what this means about the culture
Adam: Okay
T: What you think it means about the culture. And then you need to refer to your sources and say..you know.. in India traditionally the relationships between mothers and sons are like this.
Adam: Okay
T: That is coming from your outside sources {Adam nods} supporting the illustrations that you gave {Adam nods} and then the point that you made.
So this is support for both of those things.
Adam: A-hm
T: Okay?
Adam: Yeah
T: But if you're not specific, if you're not direct..then it can be..it may not give the desired effect
Adam: Okay
{Brief silence}

T: What else? Anything else?

Adam: {Shakes his head}

(16) Ted (NNS in IS): group conference with teacher and Dan (NS)
Teacher: {To Dan} Did you read Ted's essay?
Dan: Yeah.

T: {To Ted} Did you read Dan's essay?

Ted: {Nods}

T: We'll grill the essays. Who wants to go first?

Dan: Ted's essay is good, it's great.

T: Come on that's not fun [laughs].

Ss: {Smile}

T: {To Ted} What did you think of Dan's organization and content?

Ted: OK....Very nice.
[Brief silence]

T: {To Ted} Did you like his points?

Ted: [Silent]

T: Did you find them?

Ted: [Silent, looking at essay)]

T: Sometimes I couldn't find them.

[Brief silence]

T: {To Ted} What do you think?

Ted: OK. Better than me.

Data Source 2: Questionnaire responses
In her questionnaire, Rosy summarized her perspective on appropriate student conduct in the classroom like this: "Chinese thought: Respect. Teacher always right. If you have questions, ask personally or ask after class. Never interrupt or make teacher embarrassing." In the questionnaires, almost all of the students who responded "strongly agree" or "agree" to the statement, "The cultural norms I grew up with prevent me from being critical of my classmates' work in peer revision", were Chinese.

Data Source 3: Interview Information
In our one-on-one interviews Lee admitted that "[Chinese students] don't really talk, they are quiet. They aren't active," and in his interview Adam explained this behavior by stating that "Chinese students aren't comfortable criticizing other students' essays. It is rude. Criticizing brings a lot of shame. In Hong Kong, I never criticized others so I don't know. Criticism from teacher is OK, that is what happens in Hong Kong...In Hong Kong whatever teacher say is right. No questioning the teacher."

7. *Interpretation:* The video recordings present the Chinese students as one of the quietest groups in classroom discussions. However, they seem to be collaborative and even talkative in small group tasks. Adam, for example, doesn't speak up much in class, but he seems really engaged in small group work. He has smart comments and helpful ideas while performing the task, but he is never the group spokesperson. An-

other example is Paige and Jeanie who did a class presentation together. They were soft-spoken and hesitant while addressing the whole class, but they are more active in group work. Looking at the examples from various data sources and the comments of the Chinese students, it is possible to confirm Chinese students' behavior as consistently respectful rather than passive. By their own report, they consider teachers' suggestions and comments to be essential, so they listen without interrupting or asking questions, which they say is a result of their cultural background and the Chinese education system. Their discomfort in criticizing others is confirmed through group conference behavior and interview responses.

8. *Literature Comparison:* All of these results were confirmed by comparison with the literature on Chinese education, which shows that Chinese classrooms are traditionally teacher-oriented; students don't ask questions or work on collaborative tasks like small group or dyadic activities (Su & Su 1994; Liu 2001; Hammond & Gao 2002). The teacher is seen as the authority figure who lectures for the most part and doesn't encourage students to ask questions or express their opinions; students are expected to stay quiet without asking questions or challenging the teacher (Su & Su 1994). Two Asian participants in Liu's (2001) study suggested that asking very simple and basic questions in class was wasting class time. Therefore, they preferred to talk to the teacher and ask questions after class. They were surprised and annoyed to see Americans asking every question they had during class time and thereby taking up the teacher's lecture time.
9. *Conclusion:* Through triangulation of the various data sources as well as confirmation from the literature about Chinese education and other studies of Chinese students in American classrooms, it is possible to interpret Chinese students' quiet behavior and their reluctance to do peer review as related to Chinese cultural and educational norms – which are then transferred to the American setting.

   Thus, we have seen that triangulation is challenging but worthwhile; it enables researchers to look at speech communities from various angles and helps them to come up with relevant, objective, and trustworthy results.

5. Conclusion: Discourse and cognition

In this chapter on the study of discourse, we have shown the importance of working with a number of different methodological approaches, for all of which it is necessary to have a corpus of tape- and/or video-recorded language use produced within a particular socio-cultural context that can serve as the empirical data for a rich, complex analysis. In addition to the corpora themselves, the researchers collected ethnographic and sociolinguistic information through interviews, questionnaires, and/or observation, in order to become familiar with the cultural context within which the data was collected and to understand the participants themselves, including their conceptualization of the language they used and their style of interaction. They also performed fine-grained qualitative and/or quantitative analyses of discursive and interactional patterns by working with the transcripts of the language use in combination with the ethnographic and sociolinguistic information. In addition, the sources of data were triangulated to obtain rich, multiplicitous analyses. The use of a corpus of authentic language enabled the researchers not only to address their initial questions but also to explore those that emerged from the data themselves during analysis. The discussion of the corpora collected by Waugh, Fonseca-Greber, Vickers, and Eröz demonstrate that the study purpose drives the specific data collection method(s), and that the data drives the specific analytic approach(es) so that the end result is a rich analysis that captures the true complexity of human interaction in specific speech communities and communities of practice.

As said in the introduction, it is our belief that discourse and cognition are intertwined with each other and thus that the study of one of these facets necessarily touches on the study of the other: (individual) cognition arises out of, and is influenced by, language use in its multifarious social and cultural contexts. Further, cognitive representations speakers have of their language may evolve over time through a combination of factors, e.g., language change and socio-cultural change. Discourse, cognition, social and cultural activities – all are mutually implicating, and none is understandable without serious consideration of their interdependent integration. In other words, we agree with Tomasello that language itself is a "complex mosaic of cognitive and social communicative activities closely integrated with the rest of human psychology" (1998: ix). Hence, in order to be complete, the study of cognition needs to be correlated with work on discourse, social and cultural structures, and the functional foundation of language. Thus, work on discourse, especially insofar as it is multidisciplinary and integrates perspectives that take into account not only its linguistic but also its social and cultural aspects, provides a complementary, indeed a necessary, perspective for an integrated understanding of cognition. As discourse analysts, we are interested in the nature of human interaction through language in its social and cultural setting, and it is from that perspective that we can draw conclusions about, and provide insight into, cognition. Thus, as we have seen, the grammatical structures that participants use index social status, identity, attitudes, ideologies; the particular strategies chosen in the process of interaction construct hierarchy among individuals and hence create group dynamics; and the linguistic, cultural and cognitive patterns that have been developed in one culture may be difficult to change in another one and may give rise to specific behaviors that are inappropriate or misunderstood in a different cultural context. The social and cultural settings and the communicative tasks at hand influence not only what is said but also the construction, co-construction, and re-construction of the ways that individuals are perceived by others, how they perceive themselves and how groups of people interact. Thus, as said earlier, the main thinking underlying this chapter is that language use (discourse), language structure, cognition, social and cultural structures, and biological, historical and ideological patterns are all interconnected in a complex ecological system, which therefore necessitates multiple empirical approaches that can give an understanding of their relational and interdependent complexity.

References

Anderson, S. & Keenan, E. (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description III: Grammatical categories and the lexicon* (pp. 259–308). Cambridge: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Brown, R. & Gilman, A. (1960). The pronouns of power and solidarity. In T. Sebeok (Ed.), *Style in language* (pp. 253–276). Cambridge, MA: MIT Press.

Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Bybee, J. (2003). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 145–167). Mahwah, NJ: Erlbaum.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar.* Chicago: Chicago University Press.

Bybee, J. & Fleischman, S. (Eds.). (1995). *Modality in grammar and discourse*. Amsterdam: John Benjamins.

Chafe, W. (1994). *Discourse, consciousness, and time*. Chicago: University of Chicago Press.

Chafe, W. & Nichols, J. (1986). *Evidentiality: The linguistic aspect of epistemology*. Norwood, NJ: Ablex.

Chambers, J. (1995). *Sociolinguistic Theory: Linguistic Variation and its Social Significance.* Oxford: Blackwell.

Channell, J. (1994). *Vague language.* Oxford: Oxford University Press.

Cheshire, J. & Trudgill, P. (1998). *The sociolinguistics reader: Volume 2, gender and discourse.* New York, NY: Arnold.

Comrie, Bernard (1976). *Aspect*. Cambridge: Cambridge University Press.

Conrad, S. (2002). Corpus linguistics approaches for discourse analysis. In M. McGroarty (Ed.), *Annual review of applied linguistics, 22*, *Discourse and dialogue* (pp. 32–51). Cambridge: Cambridge University Press.

Coulthard, M. (1977/1985). *An introduction to discourse analysis.* London: Longman.

Coupland, N. & Jaworski, A. (1997). *Sociolinguistics: A reader and coursebook.* New York: St. Martin's Press.

Coupland, N., Sarangi, S., & Candlin, C. (2001). *Sociolinguistics and social theory*. Essex, England: Pearson.

Downes, W. (1984/1998). *Language and society*. Cambridge: Cambridge University Press.

Du Bois, J. W. (2003). Discourse and grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 47–87). Mahwah, NJ: Erlbaum.

Duranti, A. (1997). *Linguistic anthropology.* Cambridge: Cambridge University Press.

Duranti, A. & Goodwin, C. (Eds.). (1992). *Rethinking context: Language as an interactive phenomenon.* Cambridge: Cambridge University Press.

Eckert, P. (2000). *Linguistic variation as social practice*. Oxford: Blackwell.

Edwards, D. (1997). *Discourse and cognition*. London: Sage.

Edwards, J. (1993). Principles and contrasting systems of discourse transcription. In J. Edwards & M. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 3–32). Hillsdale, NJ: Erlbaum.

Edwards, J. (2001). The transcription of discourse. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 321–348). Malden, MA: Blackwell.

Edwards, J. & Lampert, M. (Eds.). (1993). *Talking data: transcription and coding in discourse research.* Hillsdale, NJ: Erlbaum.

Eröz, B. (2003). *An Ethnographic examination of international students in English composition classes: Cultural patterns, classroom dynamics, and adjustment difficulties.* Unpublished doctoral dissertation. Tucson, AZ: University of Arizona.

Fairclough, N. L. (1995). *Critical discourse analysis: The critical study of language*. Harlow, UK: Longman.

Fairclough, N. L. & Wodak, R. (1997). Critical discourse analysis. In T. van Dijk (Ed.), *Discourse as structure and process. Discourse studies: A multidisciplinary introduction* (pp. 258–284). London/Thousand Oaks/New Delhi: Sage Publications.

Fasold, R. (1984). *Sociolinguistics of society*. Cambridge, MA: Blackwell.

Fasold, R. (1990). *Sociolinguistics of language*. Cambridge, MA: Blackwell.

Fonseca-Greber, B. (2000). *The change from pronoun to clitic to prefix and the rise of null subjects in spoken Swiss French*. Unpublished doctoral dissertation. Tucson, AZ: University of Arizona.

Fonseca-Greber, B. & Waugh, L. R. (2003a). On the radical difference between the subject personal pronouns in written and spoken European French. In P. Leistyna & C. F. Meyer (Eds.), *Corpus analysis: Language structure and language use* (pp. 225–240). Amsterdam/New York: Rodopi.

Fonseca-Greber, B. & Waugh, L. R. (2003b). The subject clitics of European conversational French: Morphologization, grammatical change, semantic change, and change in progress. In R. Núñez-Cedeno, L. López, & R. Cameron (Eds.), *A Romance Perspective on Language Knowledge and Use* (pp. 99–118). Amsterdam & Philadelphia, PA: John Benjamins.

Fox, B. (Ed.). (1996). *Studies in anaphora*. Amsterdam: John Benjamins.

Georgakopoulou, A. & Goutsos, D. (1997). *Discourse analysis.* Edinburgh: Edinburgh University Press.

Givón, T. (1979). *Syntax and semantics: Volume 12, discourse and syntax.* New York: Academic Press.

Givón, T. (1989). *Mind, code, and context*: *Essays in pragmatics*. Hillsdale, NJ: Erlbaum.

Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Goodwin, C. & Heritage, J. (1990). Conversation Analysis. *Annual Review of Anthropology, 19*, 283–307.

Grundy, P. (1995). *Doing Pragmatics.* London: Arnold.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Gumperz, J. J. (2001). Interactional sociolinguistics: A personal perspective. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 215–228). Malden, MA/Oxford, UK: Blackwell.

Gumperz, J. & Cook-Gumperz, J. (1982). Introduction: language and the communication of social identity. In J. Gumperz (Ed.), *Language and social identity* (pp. 1–21). Cambridge: Cambridge University Press.

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Halliday, M. A. K. (1978). *Language as social semiotic*. Baltimore: University Park Press.

Hammond, S. C. & Gao, H. (2002). Pan Gu's paradigm: Chinese education's return to holistic communication in learning. In X. Lu, W. Jia, & D. R. Heisey (Eds.), *Chinese Communication Studies. Contexts and Comparisons* (pp. 227–243). Westport, CT: Ablex Publishing.

Heine, B., Claudi, U., & Hunnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. Chicago: University of Chicago Press.

Hill, J. & Irvine, J. (1992). *Responsibility and evidence in oral discourse*. Cambridge: Cambridge University Press.

Holmes, J. (1992/2001). *Introduction to Sociolinguistics*. Essex, Eng: Longman.

Hopper, P. (1982). Aspect between discourse and grammar: An introductory essay for the volume. In P. Hopper (Ed.), *Tense-aspect: Between semantics and pragmatics* (pp. 3–18). Amsterdam: John Benjamins.

Hopper, P. & Traugott, E. (1993). *Grammaticalization.* Cambridge: Cambridge University Press.

Hudson, R. (1980/1996). *Sociolinguistics.* Cambridge: Cambridge University Press.

Hutchby, I. & Wooffitt, R. (1998). *Conversation analysis: Principles, practices, and applications*. Cornwell: Polity Press.

Hymes, D. (1974). *Foundations in sociolinguistics*. Philadelphia: University of Pennsylvania Press.

Jacoby, S. & Ochs, E. E. (1995). Co-construction: An introduction. *Research on language and social interaction, 28*(3), 171–183.

Jakobson, R. (1936/1982). Contribution to the general theory of case: General meanings of the Russian cases. In L. Waugh & M. Halle (Eds.). (1982), Roman Jakobson. *Russian and Slavic grammar: Studies 1931–1981.* Berlin: Mouton.

Jakobson, R. (1960). Linguistics and poetics. In T. Sebeok (Ed.), *Style in language*. Cambridge, MA: MIT Press.

Jaworski, A. & Coupland, N. (Eds.). (1999). *The discourse reader*. London/New York: Routledge.

Johnstone, B. (2002). *Discourse analysis*. Oxford: Blackwell.

Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Addison, Wesley, Longman.

Klein-Andreu, F. (Ed.). (1983). *Discourse perspectives on syntax*. New York: Academic Press.

Kramsch, C. (2002). 'How can we tell the dancer form the dance?' In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (1–30). London/New York: Continuum.

Labov, W. & Waletzky, J. (1967). Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on verbal and visual arts* (pp. 12–44). Seattle: University of Washington Press.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, W. (2001). *Principles of linguistic change: Volume 2, Social factors*. Malden, MA: Blackwell.

Lantolf, J. P. (2000) Introducing sociocultural theory. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Lazarton, A. (2002). Quantitative and qualitative approaches to discourse analysis. In M. McGroarty (Ed.), *Annual Review of Applied Linguistics, 22, Discourse and dialogue* (pp. 32–51). Cambridge: Cambridge University Press.

Lee, D. (2001). *Cognitive linguistics: An introduction*. Oxford: Oxford University Press.

Levinson, S. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Liu, J. (2001). *Asian students' classroom interaction patterns in U.S. universities: An emic perspective.* Westport, CT: Ablex Publishing.

Lyons, J. (1977). *Semantics 2.* Cambridge: Cambridge University Press.

Markee, N. (2000). *Conversation analysis.* Mahwah, NJ: Erlbaum.

McGroarty, M. (Ed.). (2002). *Annual review of applied linguistics 22: Discourse and dialogue*. Cambridge: Cambridge University Press.

Moerman, M. (1988). *Talking culture: Ethnography and conversation analysis*. Philadelphia, PA: University of Philadelphia Press.

Mühlhäusler, P. & Harré, R. (1990). *Pronouns and people: The linguistic construction of social and personal identity*. Cambridge, MA: Blackwell.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Scheiffelin (Eds.), *Developmental pragmatics*. New York: Academic Press.

Peirce, C. S. (1878/1992). Deduction, induction, and hypothesis. In N. Houser & C. Kloesel (Eds.) (1992), *The essential Peirce: Selected philosophical writings. Volume 1* (*1867–1893*). Bloomington, IN: Indiana University Press.

Phillips, N. & Hardy, C. (2002). *Discourse analysis: Investigating processes of social construction*. Thousand Oaks, CA: Sage.

Saville-Troike, M. (1982/1989/2003). *The ethnography of communication*. London: Blackwell.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schiffrin, D. (1994). *Approaches to Discourse*. Oxford: Blackwell.

Schiffrin, D., Tannen, D., & Hamilton, H. E. (Eds.). (2001). *The handbook of discourse analysis*. Malden, MA/Oxford, UK: Blackwell.

Schiffrin, Deborah (2001). Discourse markers: Language, meaning, and context. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 54–75). Malden, MA/Oxford, UK: Blackwell.

Scollon, R. & Scollon, S. W. (1995/2001). *Intercultural communication.* Malden, MA: Blackwell.

Stubbs, M. (1996). *Text and corpus analysis. Computer-assisted studies of language and culture*. Oxford, UK/Cambridge/MA: Blackwell.

Su, Z. & Su, J. (1994). Teaching and learning science in American Chinese high schools: A comparative study. *Comparative Education, 30*(3), 255–271.

Tannen, D. (1984). *Conversational style: Analyzing talk among friends*. Norwood, NJ: Ablex.

Tannen, D. (1989). *Talking voices: Repetition, dialogue, and imagery in conversational discourse.* Cambridge: Cambridge University Press.

Tannen, D. (Ed.). (1993). *Framing in discourse*. New York: Oxford University Press.

Taylor, J. (1989/1995). *Linguistic categorization: Prototypes in linguistic theory.* Oxford: Oxford University Press.

Thomason, S. (2001). *Language contact: An introduction*. Washington, DC: Georgetown University Press.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam/Philadelphia: John Benjamins.

Tomasello, M. (1998). Introduction: A cognitive-functional perspective on language structure. In Tomasello 1998 (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. vii–xxiii). Mahwah, NJ: Erlbaum.

Tomasello, M. (Ed.). (1998, 2003). *The new psychology of language: Cognitive and functional approaches to language structure,* Vols. 1 & 2. Mahwah, NJ: Erlbaum.

Ungerer, F. & Schmid, H. (1996). *An introduction to cognitive linguistics*. London: Longman.

van Dijk, T. A. (Ed.). (1997a). *Discourse as structure and process. Discourse studies: A multidisciplinary introduction.* Volume 1. London/Thousand Oaks: Sage.

van Dijk, T. A. (Ed.). (1997b). *Discourse as social interaction. Discourse studies: A multidisciplinary introduction.* Volume 2. London/Thousand Oaks: Sage.

van Dijk, T. A. (1998). *Ideology: A multidisciplinary study.* London: Sage.

van Dijk, T. A. (2001a). *Critical discourse analysis*. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp. 352–371). Malden, MA/Oxford, UK: Blackwell.

van Dijk, T. A. (2001b). Multidisciplinary CDA: A plea for diversity. In R. Wodak & M. Meyer (Eds.), (2001), *Methods of critical discourse analysis* (pp. 95–120). London: Sage.

van Lier, L. (2000). From input to affordance: Social-interactive learning from an ecological perspective. In J. P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 245–259). Oxford: Oxford University Press.

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Boston/Dordrecht: Kluwer Academic.

Verschueren, J. (1999). *Understanding pragmatics*. London: Arnold.

Vickers, C. (2004). *Interactional accommodation and the construction of social roles among culturally diverse undergraduates*. Unpublished doctoral dissertation. Tucson, AZ: University of Arizona.

Virtanen, T. (Ed.). (2004). *Approaches to cognition through text and discourse*. Berlin: Mouton de Gruyter.

Wales, K. (1996). *Personal pronouns in present-day English.* Cambridge: Cambridge University Press.

Watts, R. (2003). *Politeness*. Cambridge: Cambridge University Press.

Waugh, L. R. (1982). Marked and unmarked – a choice between unequals in semiotic structure. *Semiotica, 38, 3/4*, 299–318.

Wodak, R. & Meyer, M. (2001). *Methods of critical discourse analysis.* London: Sage.

Wood, L. & Kroger, R. (2000). *Doing discourse analysis: Methods for studying action in talk and text.* London: Sage.

Yule, G. (1996). *Pragmatics*. Oxford University Press.

A case for a Cognitive corpus linguistics

Stefan Grondelaers, Dirk Geeraerts and Dirk Speelman

1. Introduction

The methodological position of Cognitive Linguistics vis-à-vis corpus research (and the use of empirical methods at large, for that matter) seems to be characterized by a certain amount of reluctance. On the one hand, corpus usage was present in Cognitive Linguistics from its inception onward. The methodology of European studies in Cognitive Linguistics in particular has tended to be more corpus-based than the early American studies, which were predominantly introspective. Whereas the use of corpus materials seems to have come to the attention of the broader community of Cognitive Linguistics only since Kemmer & Barlow (2000), it was already part of early European studies like Rudzka-Ostyn (1988), Goossens (1990), Dirven & Taylor (1988), Schulze (1988), Geeraerts, Grondelaers & Bakema (1994).

On the other hand, however, the type of quantitatively well-founded corpus-based investigations that may be found in the work of Gries (2003), Stefanowitsch (2003), Stefanowitsch & Gries (2003), Grondelaers (2000), Grondelaers, Speelman & Geeraerts (2002), Speelman, Grondelaers & Geeraerts (2003) is still rather exceptional.

While the reasons for this relative lack of enthusiasm may to some extent be practical, one cannot exclude the possibility of a more principled rejection: Cognitive Linguistics considers itself to be a non-objectivist theory of language, whereas the use of corpus materials involves an attempt to maximalize the objective basis of linguistic descriptions. Is an objectivist methodology compatible with a non-objectivist theory? We will argue presently that the contradiction is misguided, but let us note first that there are compelling reasons anyway for Cognitive Linguistics to embrace corpus research.

First, there is a growing tendency in Cognitive Linguistics to stress its essential nature as a usage-based linguistics: the central notions of usage-based linguistics have been programmatically outlined in different publications (Langacker 1990; Kemmer & Barlow 2000; Bybee & Hopper 2001b; Croft & Cruse 2004; Tomasello 2000), and a number of recent volumes show how the programme can be put into practice (Barlow & Kemmer 2000; Bybee & Hopper 2001a; Verhagen & Van de Weijer 2003). The link between the self-awareness of Cognitive Linguistics as a usage-based form of linguistic investigation and the deployment of empirical methods is straightforward: you cannot have a usage-based linguistics unless you study actual usage – as it appears in corpora in the form of spontaneous, non-elicited language data, or as it appears in an on-line and elicited form in experimental settings.

Second, the very emphasis that Cognitive Linguistics places on the fact that our knowledge of the world is an active construal rather than a passive reflection of an objectively given world, favours an interest in differences of construal between cultures, social groups, or even individuals. But such a variational perspective cannot, by definition, be realized individually. If we assume that language is not genetically so constrained as to be uniform all over the globe, and that linguistic communities are not homogeneous (two assumptions that would seem to be congenial to the non-objectivist stance of Cognitive Linguistics), then a broader empirical basis than one's own language use is necessary to study the variation.

So, how do these principled reasons for corpus research relate to the principled objection mentioned above? The objection seems to rest on a category mistake. It is not because the cognitive processes that we describe on the object level are non-objectivist that our description on the theoretical meta-level will necessarily be unable to achieve objectivity. The cognitive phenomena that you study may be non-objectivist, but you do try to study them objectively. That's what psychology does succesfully, doesn't it? And even if a non-objectivist methodology is residually unavoidable (see Geeraerts 1985 for the intricacies of the relationship between epistemological levels and meta-levels), general principles of repeatability and comparability of results favour an empirical approach spiring towards maximal objectivity.

It follows, to be sure, that corpus research is neither automatic, nor necessarily free from the hermeneutic, interpretative features that are typical of a non-objectivist methodology ("we have to understand what people mean when they mean, and such an understanding requires interpretation"). Corpus research, in fact, neither denies nor ignores the necessity of interpretations, but it takes on a helix-like structure of a gradual refinement of interpretations through a repeated confrontation with empirical data. An initial hypothesis, which will more often than not be derived introspectively, is confronted with the corpus data; interpreting the results leads to a more refined hypothesis and more questions, which may then be subjected to further testing – and so on.

Crucially, then, corpus research rests on the *operationalization of hypotheses*: if we interpretatively assume that a linguistic entity has this meaning or that function, what can we expect about its observable behaviour in actual language use? The further steps for a corpus approach follow logically: how can we identify that behaviour in the corpus data, and when can we say that the initial hypothesis is correct? These two further steps involve the instruments of corpus research that tend to get the most attention in introductions and textbooks: corpus design and corpus tools for data retrieval, and statistics for testing the data. From a methodological point of view, however, the preliminary step (even though it is less often discussed) is the really essential one: linking interpretative hypotheses to observable corpus phenomena – and repeating the procedure in a process of gradual refinement of the hypotheses.

In the remainder of this paper, we will illustrate the added value that a corpus investigation may bring on the basis of a case study involving the intriguing Dutch particle *er* "there", as used in presentative sentences like *Er staan twee schoorstenen op het dak* "there are two chimneys on the roof". We will exemplify the helix-like nature of empirical research, and we will demonstrate how corpus research and experimental investigations may be complementary.

2. The case

Few linguistic phenomena have given rise to such fiery theoretical controversy as presentative sentences and, in particular, presentative *there* (a cognitive approach to *there* can be found in Lakoff 1987; Kirsner 1979 is an early but excellent example of a proto-cognitive analysis of Dutch *er*). Examples (1)–(3) illustrate the standard type of presentative sentence in Dutch, which is to a high extent isomorphic with English *there*-sentences. The sentence-initial *er* in this construction type has been dubbed "plaatsonderwerp" (topical subject), "repletive *er*", "presentative *er*" (Haeseryn et al. 1997: 464, 467ff.), and "existential *er*":

(1) *Er* bevonden zich geen paparazzi *vlak voor, naast of achter de wagen.*
There were no paparazzi in front of, next to, or behind the car.

(2) *Er* zijn op dit moment 11.500 invalide zelfstandigen.
There are at this moment 11.500 disabled businessmen
"There are 11.500 disabled businessmen at this moment."

This case study, however, concentrates on the distribution of *er* in the adjunct-initial presentative sentence type. When an adjunct is topicalized in Dutch presentatives, *er* emerges postverbally, if it emerges at all: it has often been observed that *er* can in certain cases be deleted in adjunct-initial presentative sentences (see Haeseryn et al. 1997; and De Rooij 1991 for Dutch, and Bolinger 1977 for English).

(3) In 1977 was *er* een fusie tussen Materne en Conflux.
In 1977 was *er* a merger between Materne and Conflux
"In 1977 there followed a merger between Materne and Conflux."

(4) Morgen volgt een extra ministerraad.
Tomorrow follows an additional cabinet meeting
"Tomorrow there is an additional cabinet meeting."

(5) In het redactielokaal staan enkele flessen wijn en wat borrelhapjes.
In the newsroom stand some bottles of wine and some appetizers
"In the newsroom there are some bottles of wine and some appetizers."

(6) In ons land is *er* nog altijd geen openbaar golfterrein.
In our country is *er* still no public golf course.
"In our country there still is no public golf course."

The restriction to adjunct sentences has the pivotal methodological advantage – first observed in Bolinger's thought-provoking study on the distribution of *there* in English adjunct-initial presentatives (1977: 93) – that any syntactic, semantic or functional difference between sentence variants with and without *er* can only be attributed to the presence or absence of *er*. Conversely, by identifying such syntactic, semantic and functional differences between sentence variants with and without *er*, one can "tease out the true sense of" *er* (idem).

The latter, unfortunately, has turned out to be a frustrating enterprise: *er*'s postverbal distribution has given rise to a good deal of descriptive vagueness in the linguistic literature. According to the Algemene Nederlandse Spraakkunst – the standard grammar of Dutch (Haeseryn et al. 1997: 473) – no strict rules can be given for the presence or absence of postverbal *er*: "it can be optional, there may be semantic or stylistic differences involved, and there is a lot of individual and sometimes also regional variation in its use".

Only a handful of factors have hitherto been identified (mostly on the basis of introspection). On p. 477, specifically in connection with the optional character of postverbal *er* in adjunct-initial sentences, the ANS states: "in the standard language, *er* is more easily deleted in sentences with a fronted locative than in other cases" (see also Van Es & Van Caspel 1971). Another observation (made, for instance, in De Rooij 1991) which converges well with native speakers' intuitions is the idea that the preference for *er* correlates negatively with the taxonomical specificity of the main verb: the more specific the verb, the less *er*. To complicate matters further, there are outspoken regional and stylistic tendencies in *er*'s postverbal distribution: Belgian Dutch is known to be much more tolerant towards *er* than Netherlandic Dutch (cf. De Rooij 1991; Haeseryn et al. 1997), and *er* is attested significantly more often in *informal* register (Haeseryn et al. 1997).

3. Preliminary analysis

3.1 The evidence revisited: A regression analysis of four factors

Confronted with such distributional complexity and scant evidence, how can we proceed to investigate *er*'s distribution in an empirically more responsible way? In the next section we will outline and operationalize a functional hypothesis which will subsequently be verified on the basis of an extensive corpus of written Dutch. Let us first, however, reconsider the four factors in a quantitatively more dependable way, in order to be able to feed them into the hypotheses. For now, it is essential that we find a reliable answer to the following questions. First, are the four factors really significant determinants of *er*'s distribution? All of them are compatible with speaker judgements, but none of them constitute clear-cut rules: (6), for instance, is an obvious counterexample to the "rule" that locative adjunct sentences are preferably constructed without *er*. Second, what is their respective impact on *er*'s distribution? This question is especially relevant in this context, since the factors involved come in two types: whereas ADJUNCT TYPE (temporal or locative) and VERBAL SPECIFICITY are *language-structural factors* which – when succesfully identified and operationalized – help solve *er*'s distribution, REGION and REGISTER are *contextual factors*, which traditionally belong to the domain of sociolinguistics, and which constrain our explanations to one region or register! More specifically, when we find that ADJUNCT TYPE *does* have an impact, we are one step closer to our goal, viz. determining *er*'s distribution; if, by contrast, the statistical analysis reveals that, for instance, REGION has a significant influence on *er*, we know that our account is restricted to Belgian or Netherlandic Dutch, which greatly complicates matters. It is therefore pivotal that we learn which factors are significant, and which factors – language-structural or contextual – have the greatest impact. Third, how much variation can be explained and predicted on the basis of these four factors? Are they already sufficiently powerful predictors, or do we need additional features?

A reliable answer to these questions involves four steps which will be dealt with in the following subsections: (i) the compilation of a representative corpus (3.2), (ii) the implementation of the dependent variable – the variation pattern we wish to measure (viz. the preference for *er* in adjunct-initial presentative sentences) – and the independent variables (the four factors believed to influence the preference for *er*) (3.3), (iii) the calculation of the raw frequencies of the dependent variable in the conditions defined by the independent variables, and (iv) a regression analysis to determine the significance, the respective impact, and the predictive success of the independent variables (3.4).

3.2 Materials

Our variationist approach to *er*'s distribution necessitates a sophisticated corpus design. Notice first and foremost that our suspicion that *er*'s distribution is affected by contextual factors requires the presence in the corpus of both Belgian and Netherlandic Dutch, as well as formal and informal materials. In the absence of a fully stratified corpus of Dutch, we compiled the ConDiv-corpus, an extensive text-database tailored to the needs of various corpus-based sociolinguistic investigations (see Grondelaers et al. 2000 for design details).

The ConDiv-corpus not only contains Netherlandic as well as Belgian Dutch texts, it is also structured along an important register dimension. The corpus basically consists of two types of attested language use. In the newspaper component a distinction is made between quality newspapers such as *De Standaard* and *NRC Handelsblad*, and popular newspapers such as *De Telegraaf* and *Het Laatste Nieuws*. There are national popular papers – *Het Laatste Nieuws en De Telegraaf* –, but also regional popular newspapers such as *Het Belang van Limburg* or *De Gazet van Antwerpen*.

In addition to newspaper language, the ConDiv-corpus also contains more informal language data attested on the Internet. From Geeraerts, Grondelaers & Speelman (1999), we know that it is incorrect to conceive of Belgian Dutch as a monostratal language. In-between standard language and dialects, there is at least one intermediate level on which a higher degree of informality coincides with geographical specialization: the more informal the communication setting, the more regional the Dutch sounds in which (especially Belgian) speakers express themselves. In recent publications this intermediate register has been dubbed "tussentaal" (Taeldeman 1992: 33–52), "verkavelings-Vlaams" (Van Istendael 1993: 116), or "soap-Vlaams" (Geeraerts 1999: 232).

Since it is impossible to determine a priori how many intermediate strata must be distinguished in Dutch, we did not look for language use which represents a certain stratum. Instead, we considered the different communicative situations in which written Dutch is produced as independent variables, and the language spoken in these situations as the dependent variable. Hence, the stylistic-stratificational variation in Dutch is accommodated in the corpus by comparing language data from four different communicative situations, which can be positioned on a stylistic scale between "informal" and "very formal". The three types of newspaper materials – regional popular newspapers, national popular newspapers and quality papers – occupy the highest positions on the formality scale. The lowest position on the scale – informal Dutch – is represented in this study by language data attested on UseNet, an Internet forum on which surfers debate in "newsgroups", by means of e-mail messages they add to an ongoing discussion. Since e-mail is offline – so that users can reread their contributions before adding them to a "thread" – and since academic Internet operators only tolerate (relatively) serious newsgroups on their net –, the UseNet register is not as informal as Internet Relay Chat, a module in which anonymous users debate online.

A second material prerequisite for a responsible investigation of postverbal *er*'s distribution is language materials in which *er*'s distribution is as representative as possible for its actual usage. For this reason, no Internet Relay Chat (cf. previous paragraph) was included in the analysis. The typical interactional characteristics of chat – "temporality and immediacy" (Bays 1998) – necessitate specific formulation techniques "to augment the speed and the capacity of information transfer" (idem). The most important strategies in this respect are "abbreviation, ellipsis and a telegraphic style, which reduce the quantity of words that need to be typed, sent and read" (idem). Needless to say that postverbal *er* is an endangered linguistic species in the context of this condensed style, especially there were it is not needed for grammaticality: a series of regression analyses in Grondelaers (2000: 193–196) confirmed that the telegraphic style which is characteristic of chat has a detrimental effect on the use of *er* in that medium. As a result, we restrict the analysis to language materials in which production speed plays no role.

Production speed, however, is not the only menace to a representative *er*-distribution. Corpus sceptics like Verkuyl (1998: 63ff.) also mention "proof reader-idiosyncrasies": postverbal *er* is particularly vulnerable to the last minute interferences of press revisers or editors-in-chief. Some papers, in addition, are known to have style guides which are excessively hostile to the "unnecessary" postverbal use of *er*. To reduce the impact of such hostility, we have restricted the analysis to language data from newspapers which, when asked about their attitude towards *er*, emphatically stated not to have any such policy. None of the Belgian newspapers turned out to devote special attention to *er*; *De Telegraaf* and *NRC* were the only Dutch newspapers which responded to our question, and they too assured us that *er* is in no way stigmatized in their publications.

3.3 Dependent and independent variables

From the reduced ConDiv-corpus, we extracted all inverted presentative main sentences constructed either with a locative or a temporal adjunct. Observations like *Bij die ramp vielen 34 doden* "In this disaster 34 people were killed" demonstrate that the opposition locative vs. temporal is not a binary distinction, because *bij die ramp* allows at the same time a locative ("on the place of the disaster") and a temporal interpretation ("at the time of the disaster"). Because of its low frequency, observations with this intermediary adjunct type were excluded from the analysis. For reasons which will become clear, we are primarily interested in constructions with one overt subject, as a result of which imper-

Table 1. Distribution of observations over the sources in the corpus

	UseNet	popular papers		quality papers
		De Telegraaf		*NRC*
N	n = 192	n = 227		n = 263
		Het Belang van Limburg	*Het Laatste Nieuws*	*De Standaard*
B	n = 225	n = 397	n = 198	n = 403

sonal passives such as *Er werd gedanst* "there was dancing" – which need not contain such a subject in Dutch – were also barred. The distribution of the extracted observations over the different sources in the corpus is given in Table 1 above.

The preference for *er* was subsequently quantified as the ratio between the absolute frequency of the adjunct-initial clauses attested with *er* in a source, and the total frequency of adjunct-initial clauses in that source. This calculation returns the maximal *er*-preference value for adjunct-initial clause types which always contain *er* and the minimal value for clauses which are never attested with *er*.

The implementation of the contextual independent variables is unproblematic in this analysis, for REGION – Belgian vs. Netherlandic Dutch – and REGISTER – UseNet vs. popular newspapers vs. quality papers – are reflected in the structure of the corpus. The extracted observations were then tagged for the language-structural variables ADJUNCT TYPE and VERBAL SPECIFICITY.

ADJUNCT TYPE was operationalized straightforwardly by contrasting observations with a locative and a temporal adjunct. The verbal specificity factor, by contrast, does not translate easily into an operational parameter. Although adjunct-initial presentative sentences allow only a limited number of verbal classes, the specificity of those verbs may be determined by any of the three conceptual ingredients their semantics presuppose: nearly all verbs in adjunct-initial presentative sentences such as *On the roof was a bird* code a relation between the referent of the subject ("a bird") and the temporal or locative setting the adjunct refers to ("the roof"). Because of the conceptual inseparability of a verbal process from its subject and its setting, we have operationalized the verbal specificity-factor on the basis of the size of the class of possible subjects the different verbs in adjunct-initial presentatives subcategorize. Building on this criterion, the lowest level of specificity is represented by the verb *to be*, which imposes no restrictions at all on process, setting and subject. In this respect it is hardly surprising that almost all the verbs we encounter in adjunct-initial presentative sentences are hyponyms of *to be*.

The highest level of specificity, by contrast, is represented by verbs which are constructed with a limited set of subjects. These include Levin's (1993: 250) VERBS OF EXISTENCE, which are "typical of certain entities". The latter can be subdivided in VERBS OF ENTITY-SPECIFIC MODES OF BEING (*vloeien* "flow", *branden* "burn"), VERBS OF MODES OF BEING INVOLVING MOTION (1993: 251) like *wapperen* "flutter", and MEANDER VERBS like *meanderen* "meander". VERBS OF SOUND EXISTENCE (1993: 252) like *echoën* "echo" restrict their subject to sound-producing entities, whereas VERBS OF GROUP EXISTENCE (1993: 253) like *zwemmen* "swim" or *dansen* "dance" typically refer to the existence of resp. fish and bees. On the highest specificity level we also find another subcategory of the verbs of existence, i.e. Levin's (1993: 255) VERBS OF SPATIAL CONFIGURATION which, since they designate "the spatial configuration of an entity with respect to some location", impose specific restrictions on relation and location and, hence, limit the class of potential subjects. Typical examples are *zitten* "sit", *staan* "stand", *liggen* "lie", and *hang* "hangen". The same goes for a second subclass of verbs on the highest specificity level, the VERBS OF APPEARANCE, DISAPPEARANCE AND OCCURRENCE (1993: 258–261) which refer, respectively, to "the appearance of an entity on the scene" (cf. *verschijnen* "appear", *landen* "land", etc.), "the disappearance or going out of existence of some entity" (cf. *vergaan* "perish" – in contrast to English, Dutch *does* allow DISAPPEARANCE-verbs in adjunct-initial presentative sentences, cf Levin 1993: 260) – and "the occurrence of an event" (*plaatsvinden* "occur", *aan de gang zijn* "take place"). Finally, two subcategories of VERBS OF MANNER OF MOTION appear in adjunct-initial presentative sentences, dubbed the ROLL- & RUN-VERBS (1993: 264–267), although according to Levin they might just as well be considered as verbs of existence. *Drijven* "float", *glijden* "glide" and *rollen* "roll" are typical roll-verbs, whereas *rennen* "run" and *springen* "jump" are frequently attested examples of run-verbs.

Contrary to Van Es & Van Caspel (1971), De Rooij (1991), and Grondelaers & Brysbaert (1996), we do not restrict ourselves to a binary verbal specificity opposition. We add an intermediate level represented by a small group of (frequently attested) verbs which impose a minimal restriction on one of their conceptual ingredients. The verb *bestaan* "exist" is slightly more specific than *zijn* "to be" because it situates its subject within the metaphysical boundaries of this world. *Ontstaan* "to come into being" adds an inchoative aspect to zijn, *blijven* "remain" an imperfective aspect, and *voorbij gaan* "to pass" and *eindigen* "to end" a perfective aspect. The frequently attested verb *heersen* "to prevail" bestows a greater agentivity on its animate as well as inanimate subjects. So, what all these verbs have in common in addition to their somewhat schematic meaning, is the fact that they impose minimal restrictions on their subjects, without, however, becoming as unrestricted as *to be*.

3.4 Results and logistic regression analysis

The data in Table 2 were collected in order to answer three questions in connection with postverbal er. First, is the impact of the individual variables on *er's* distribution significant? Second, which factor's impact is the most outspoken? Third, is the explanatory and predictive power of the model which contains these factors as poor as traditional analyses – notably Haeseryn et al. (1997) and De Rooij (1991) – would like us to believe?

For a statistically reliable answer to these questions, we subject our data to a *logistic regression analysis*. A description of logistic regression preferably starts from a simpler variant of regression, viz. *linear regression*. The latter is based on the correlation between two sets of variable data. If, for instance, an employee gets a two-yearly increment of 2%, then his increase in earnings will correlate systematically with his seniority, and this correlation can be expressed in a formula $y = a + bx$. In this formula y is the dependent variable "salary", and x the independent variable "seniority"; a is a constant which reflects the employee's starting wage, whereas b is an estimate which reflects the degree to which the independent variable "seniority" contributes to the employee's salary at a given time. The "+" in the formula indicates that seniority has a positive impact on salary. In theory,

Table 2. Relative frequencies of *er* in adjunct-initial presentative sentences as a function of region, register, adjunct type and verbal specificity

			N			B			*avg.*
			Use	Pop	Qua	Use	Pop	Qua	*+er*
tem	*zijn*	*–er*	0.0	0.0	2.3	0.0	0.0	0.0	
		+er	100	100	97.7	100	100	100	*99.7*
	int.	*–er*	13.6	22.4	33.3	5.0	21.9	35.29	
		+er	86.4	77.6	66.7	95.0	78.1	64.7	*75.5*
	e&a	*–er*	0.0	40.0	40.0	38.1	45.5	52.0	
		+er	100	60.0	60.0	61.9	54.5	48.0	*57.6*
loc	*zijn*	*–er*	33.3	36.8	37.5	7.7	10.6	15.8	
		+er	66.7	63.2	62.5	92.3	89.4	84.2	*81.6*
	int.	*–er*	87.0	80.0	91.1	63.0	74.8	81.5	
		+er	13.0	20.0	8.9	37.0	25.2	18.5	*20*
	e&a	*–er*	97.3	100	98.4	73.9	93.8	94.9	
		+er	2.7	0.0	1.6	26.1	6.2	5.1	*6.3*
	avg + er		*48.4*	*46.7*	*38*	*66.2*	*53.3*	*38.2*	

however, it is possible that the wages of older employees are reduced by 1%, in which case the formula should be rewritten as $y = a - bx$.

A regression analysis yields two sorts of findings. On the basis of the data of all employees, the software not only computes an estimate for *b*, but also for *a*, for the latter need not be given. In our simple example, for instance, *a* can only be equated with "starting wage" in the case of employees whose salary increases; when applied to older employees (whose salary is reduced), *a* should be equated with the maximum salary an employee received. In addition, the software determines the statistical significance of *a* and *b*. It is possible that the group of employees included in the analysis is too small to be representative of the whole company, in which case the impact of seniority on salary may be no more than a coincidence. As a consequence we need an estimate of the chance that our data reflect a genuine relation. The statistical measure for the latter is the *p-value*. In linguistics it is customary to regard a p-value below 0.05 as a sufficient confirmation that a relation is genuine; in the latter case, the findings of the sample (the group of employees we investigated) can be extrapolated to the whole population (all employees in a similar situation).

If we apply this technique to a situation with more than one independent variable, we use *multiple linear regression*. Suppose that in addition to increments, occasional bonuses are included in the model. In that case, the formula will take the form of $y = a + bx1 + cx2$, whereby *x2* represents the bonus-variable. This model represents the simplest type of multiple regression analysis, with two independent variables *x1* and *x2* for which the software computes the estimates *b* and *c*, which are both tested for significance.

In linguistic research of the type illustrated in this paper, however, almost no variable can be quantified as straightforwardly as "salary" or "seniority", which can be very precisely and reliably expressed in stable figures. Our analysis makes use of "categorical" independent variables such as "main verb *to be* vs. *intermediate* main verb vs. *existence &*

Table 3. P-values and Odds Ratios of the independent variables which explain *er's* distribution in the global database

	P	O.R.
adjunct type	0.0001	40.2
verbal specificity 1	0.0001	68.8
verbal specificity 2	0.0001	3.27
region	0.0001	3.36
register 1	0.0001	2.68
register 2	0.0051	1.55

appearance main verb", which do not allow a linear correlation with the categorical dependent variable "*er* present or absent". To deal with this problem, statisticians use regression models which look for non-linear relations in the form of a logistic curve, hence the name "logistic regression" for this technique (the technical details are not important here, see Rietveld & Van Hout 1993: 330ff. for an overview).

Table 3 contains the output of the logistic regression analysis to which we subjected our data. The analysis returns *p-values* (in the second column) which reflect the statistical significance of each independent variable (as was indicated, a value below 0.05 confirms that the impact of an independent variable is significant), as well as *Odds Ratios* (in the third column) which reflect the relative importance of each factor's impact on the variation at issue. Odds Ratios are interpreted as follows: if the statistical software R – a GNU-project similar to the S-language used in the commercial packet SPSS – returns Odds Ratio "6" for an independent variable, the use of *er* vs. the non-use of *er* is predicted to increase six times as a result of the impact of that independent variable. Conversely, Odds Ratio "0.2" – which is equal of course to 1/5 – indicates that the use of *er* vs. the non-use of *er* is predicted to decrease five times as a result of the impact of that variable.

Before we go into the data in Table 3, it should be noticed that all independent variables are interpreted as nominal variables, a consequence of which is that our analysis returns two estimates for variables with three values (VERBAL SPECIFICITY and REGISTER): a first one for the impact on *er*'s distribution of the opposition between values "1" and "3" (in the case of, for instance, VERBAL SPECIFICITY, between the main verb *to be* and a main verb of the EXISTENCE & APPEARANCE-type), and a second one for the impact on *er*'s distribution of the opposition between values "2" en "3" (the difference between intermediary verbs and EXISTENCE & APPEARANCE-verbs). Again, the technical details are of no great concern.

Much more important is the confirmation in Table 3 that the impact of all factors is highly significant. For the language-structural variables ADJUNCT TYPE and VERBAL SPECIFICITY, whose impact can be readily discerned in Table 2, this outspoken statistical significance could be predicted. Slightly more surprising, however, is the high significance of both REGISTER-variables, whose effect in Table 2 is mainly restricted to the Belgian sample. And although regional variation is largely limited to the popular newspapers and, in particular, UseNet, the high significance of the variable REGION from now on necessitates extreme caution when interpreting the global data.

Despite their outspoken significance, the respective impact of these variables on *er*'s distribution differs noticeably. The fact that ADJUNCT TYPE receives an Odds Ratio of 40.2 in this model signifies that the use of *er* compared with the non-use of *er* is predicted to be more than 40 times higher in temporal adjunct sentences than in locative adjunct-sentences. The Odds Ratio 68.8 for VERBAL SPECIFICITY 1 indicates that the use of *er* vs. the non-use of *er* is predicted to increase more than 68 times when an EXISTENCE & APPEARANCE-verb is substituted with a form of *to be*. According to the Odds Ratio 3.27 for VERBAL SPECIFICITY 2, the effect of the substitution of an EXISTENCE & APPEARANCE-verb with an intermediary verb is much less far-reaching. The Odds Ratios for VERBAL SPECIFICITY indicate, in other words, that the verb *to be* is the main *er*-trigger in an adjunct-initial sentence: postverbal *er* is restricted most efficiently by using a more specific main verb than *to be*; the exact nature of the more specific verb (intermediary or EXISTENCE & APPEARANCE) is not so important according to our data.

The Odds Ratios demonstrate that the impact of the contextual variables on *er*'s behaviour is relatively limited compared to the effect of the language-structural factors. The Odds Ratio 3.36 for REGION reveals that the use of *er* vs. the non-use of *er* is predicted to be more than three times higher in Belgium than in The Netherlands. The REGISTER-variables – resp. Odds Ratios 2.68 and 1.55 – have the smallest impact on *er's* distribution.

The analysis also returns estimates of the global explanatory and predictive quality of our four factors. The fact that *er*'s distribution is correctly predicted in 84.4% of all cases contrasts sharply with the pessimistic attitude towards *er*'s distribution propounded in Haeseryn (1997: 477) and in De Rooij (1991: 127). In the absolute majority of adjunct-initial presentative sentences, *er*'s distribution can be predicted with the simplest of algorithms: neither the use of *er* in a temporal adjunct-sentence with the main verb *to be*, nor the absence of *er* in a locative adjunct-sentence with a more specific verb will ever lead to unacceptable sentences.

Judging from the noticeably lower *er*-proportions in the bottom rows of the left half of Table 2, the latter goes especially for *er*'s distribution in Netherlandic Dutch. The much higher *er*-residues in the right bottom rows reveal that *er*'s distribution is much more difficult to predict in Belgian Dutch. A separate regression analysis of the Belgian and Netherlandic data confirms this tendency. In the Netherlandic model, ADJUNCT TYPE and VERBAL SPECIFICITY are absolutely dominant: Netherlandic adjunct sentences can typically do without *er* when they have an initial locative adjunct, and the verbal factor blocks *er*-preferences in the rare cases where the locative adjunct hasn't already done so (there is no significant REGISTER-variation in the Netherlandic distribution of *er*). As a result, the predictive success rate of the Netherlandic model is as high as 89.5%. In the Belgian model, by contrast, the effect of ADJUNCT TYPE and VERBAL SPECIFICITY on *er*'s predictability is much less outspoken (locative adjunct sentences and adjunct sentences with specific main verbs are frequently attested with *er*), and the factors barely interact. In addition, REGISTER *does* play a role: the more informal the Belgian source, the more frequently adjunct sentences will contain *er*. The cumulative success rate of the three variables in Belgian adjunct sentences is no more than 85.1%.

In what follows we will therefore carry out separate analyses of *er*'s distribution in Belgian and in Netherlandic Dutch. In the next section we concentrate our efforts on finding additional factors which could further our insight into the distribution of *er* in Belgian Dutch. These additional factors no longer come from the linguistic literature on *er*, but represent (potential) corpus parameters of an encompassing functional hypothesis about presentative sentences and presentative *er*, to which we turn now.

4. Reference points and (in)accessibility markers

The basic axiom of Cognitive Linguïstics is the non-autonomous character of linguistic structure, because "fundamental cognitive abilities and experientially derived cognitive models have direct and pervasive linguistic manifestations, and conversely, (...) language structure furnishes important clues concerning basic mental phenomena" (Langacker 1993:1). In the final decades, cognitive linguists have found evidence for a number of phenomena whose relevance is not restricted to language (like Talmy's *force dynamics*, 1988, and Lakoff's *image schemas*, 1987). A recent discovery in this field was the "reference point scenario", a phenomenon so ubiquitous in our perception of our surroundings that we are largely oblivious to it. Langacker (1991:170, but see also Taylor 1996:17–19) introduces the reference point phenomenon as follows: "the world is conceived as being populated by countless objects of diverse character. The objects vary greatly in their salience to a given observer; like stars in the night-time sky, some are immediately apparent to the viewer, whereas others become apparent only if special effort is devoted in seeking them out. Salient objects serve as *reference points* for this purpose: if the viewer knows that a non-salient object lies near a salient one, he can find it by directing his attention to the latter and searching in its vicinity."

In the cognitive literature the reference point phenomenon is considered to be "the abstract basis" for semantically diverse prenominal possessives such as *the boy's watch*, *the girl's uncle*, *the dog's tail*, *the cat's fleas* or *Lincoln's assassination*. What all these constructions have in common is "that one entity (the one we call the *possessor*) is invoked as a reference point for purposes of establishing mental contact with another (the *possessed*)" (Langacker 1993:8–9). Take the example *the girl's uncle*: "the very purpose of a kinship term is to situate people – socially and genealogically – with respect to a reference individual ("ego"). Only in relation to a particular ego does it make sense to call someone a *cousin*, an *uncle*, a *sister*, or a *stepson*; a person is not a *cousin* or a an *uncle* autonomously (...)." In the same way, it only makes sense to refer to a part in the context of a whole "which functions as a natural reference point for its conception and characterization". The notion "possession" also is "clearly asymmetrical and lends itself very naturally to reference-point function. We know and recognize people as individuals, but for the most part we do not have comparable individual familiarity with their possessions (except our own). Moreover, a given person has numerous possessions, each of which he uniquely controls (according to our idealized cognitive model), and for any general type of object (e.g. watches) there are many exemplars that we know nothing about except that each belongs to a particular individual. Hence a person is naturally invoked as a "mental address" providing access to the cluster of items he possesses."

The linguistic relevance of the reference point strategy is not restricted – according to Langacker (1993: 26ff.) – to possessive constructions, because "certain presentational constructions that serve to introduce an element into the scene are reasonably attributed reference-point function" (ibidem). A case in point, Langacker continues, are the adjunct-initial clauses represented in (7)–(9):

(7) On the table sat a nervous calico cat.

(8) Beside the pond stood an enormous marble sculpture.

(9) In her room were many exquisite paintings.

Since the prepositional constituent focuses the hearer's attention on a specific location, in which a new entity is subsequently introduced (idem: 26), it is attractive to regard the preposed adjuncts in these clauses as reference points for the conception of a subject entity which is new and hence non-salient. Notice that constituent ordering in adjunct-initial presentatives "iconically diagrams the mental route that the conceptualizer needs to follow in order to identify" the located element (Taylor 1996: 18): the fact that the subject is clause-initial symbolizes that the hearer is first invited to conceptualize the adjunct as a "mental adress" at which the new entity can then be found.

The reference point scenario not only has a *conceptual* dimension to it (as sketched in the previous paragraph), but also a *discursive* dimension (elaborated for the most part in Taylor 1996: 193ff.). On the discourse level, the facilitating impact of a reference point is the result of the referential restrictions the reference point imposes on the target entity: "an effective reference point is one which will limit the choice of possible targets" (Taylor 1996: 193). The difference between target entities in possessive constructions and presentatives – reflected in the fact that possessees are typically marked definite whereas adjunct clause subjects are predominantly indefinite – boils down to the efficiency of both types of reference points: "the possessor nominal (...) delimits the referential possibility of the possessee nominal – prototypically, down to a single, uniquely identifiable entity" (Taylor 1996: 207). "In the optimal case, there will only be one possible target; the reference point is such that there is a *unique relation* between it and the intended target" (idem: 193). In presentative sentences, by contrast, there rarely is a unique relation between reference point and target: "good" reference points – we will shortly define "good" and "bad" reference points – narrow down the set of *nominal types* that can be coded by the subject NP head:

(10) a. In the refrigerator was a sausage
b. In the refrigerator was milk
c. ?In the refrigerator was bread
d. ??In the refrigerator was a toothbrush

An adjunct such as "refrigerator" typically selects its subject from the relatively small class of "food stuffs limited in size which have to be refrigerated in order to remain fresh". Although this category contains more than one member (as shown by (10a) and (10b)), its mental activation prior to its actual realization speeds up the processing of compatible target entities as in (10a) and (10b). Incompatible target entities such as *bread* in (10c) (which needs not be refrigerated) and *a toothbrush* (no food), by contrast, were shown in the psycholinguistic literature to be processed significantly slower. The psychological key concepts in this respect are *contextual constraint* – "the degree to which a sentence constrains the reader's expectations for possible completions" (Schwanenflugel 1986: 363) – and *predictive inferencing* (Altmann & Kamide 1999), the view that unfolding context functions as an incremental *filter* (updated on a constituent-by-constituent basis) which reduces the postverbal domain of reference on the basis of semantic and real-world knowledge emanating from the verb and initially realized arguments (in this case the adjunct).

The latter paragraphs bring us to a major oversight in the Cognitive Linguistics literature: Langacker does not explicitly specify the relevant linguistic characteristics of "good" reference points, except that they have "a certain cognitive salience, either intrinsic or contextually determined" (1993: 6). And Taylor (1996: 210) does not go beyond suggesting that good reference points are highly accessible, i.e. topical.

We will therefore have to operationalize the reference point phenomenon ourselves, and find quantifiable parameters of "good" reference points. Now, if good reference points enhance the processing of the subject entity by imposing restrictions on the referential potential of the subject nominal, then locative adjuncts will be better reference points than temporal adjuncts, because they situate their subjects in a spatial setting which generates more useful inferences about the upcoming subject (concerning its concreteness, size, etc.) than temporal settings, which do not constrain the entities they situate in any considerable way. Adjuncts such as *gisteren* "yesterday", or even the much more concrete *vorige week dinsdag om tien over half negen* "last tuesday at 8:40 AM", do not impose many restrictions on forthcoming subject entities, which can be abstract or concrete, of any conceivable nature or size:

(11) Vorige week dinsdag om tien over half negen was *er* een documentaire over Kenya op televisie.
"Last tuesday at 8:40 AM there was a documentary on Kenya on"

(12) Vorige week dinsdag om tien over half negen was *er* plots een zwarte vrachtwagen.
"Last tuesday at 8:40 AM there suddenly was a black lorry."

(13) Vorige week dinsdag om tien over half negen was *er* eindelijk rust.
"Last tuesday at 8:40 AM there suddenly was peace."

It should be noticed at this point that the subject reduction realized in adjunct-initial clauses is not only effected by the reference point – viz. the adjunct – itself: referential restrictions on the subject also emanate from the verb. Recall in this respect that the factor VERBAL SPECIFICITY was implemented exactly in terms of the size of the class of potential subjects the different verbs subcategorize: the verb *zijn* "to be" does not constrain its subjects in any way, intermediary verbs impose only schematic restrictions, and EXISTENCE & APPEARANCE verbs are attested exclusively with a limited class of subjects. In the light of the "extended" reference point hypothesis outlined in this paragraph – reference point facilitation is the responsibility of the adjunct *and* the verb – we suggest that the more specific a verb, the more suitable as a reference point: EXISTENCE & APPEARANCE verbs constitute better reference points than intermediary verbs, which themselves are better reference points than any form of the verb *zijn* "to be".

Let us now move on to a hypothesis about *er*'s function in adjunct-initial sentences, and observe in Table 2 that *er* is attested predominantly in the context of poor reference points, viz. temporal adjuncts and unspecific main verbs. Clauses which feature a locative adjunct and a specific main verb, by contrast, can do very well without *er* (compare the lowest row in Table 2). *Er*, in other words, appears to be a "diagnostic indicator" of poor reference point potential, of limited inferential access into the upcoming subject. We will therefore provisionally refer to it as an *inaccessibility marker* (notice that Bolinger 1977: 92 also mentions the absence of proper contextual anticipation of the subject as a *there*-triggering factor).

In (14)–(15) we propose working hypotheses on resp. the function of adjunct-initial clauses, and the function of *er* in such clauses:

(14) Adjunct sentences represent an important linguistic manifestation of our cognitive reference point potential, because they are semantically and syntactically structured to enhance inferential access into a subject entity which is new and therefore difficult to process.

(15) Adjunct sentences tend to contain *er* when one or more of their ingredients contributes insufficiently to enhancing the inferential accessibility of the subject entity.

While (14)–(15) account for the language-structural factors ADJUNCT TYPE and VERBAL SPECIFICITY in an integrated way, the latter factors do not – according to our statistical analysis – explain all the variation observed in our data. Hence, we have to look for additional factors, viz. observable parameters of *er*'s alleged function as an inaccessibility marker. Since the distribution of *er* in temporal adjunct clauses is more or less fully predictable, we will concentrate our search for extra factors on locative adjunct sentences, in which *er*'s distribution is far less predictable (as revealed in Table 2).

First, however, we have to concentrate on one of corpus linguistics' major limitations, viz. the fact that it is impossible to find corpus evidence in support of hypotheses such as (14). Phenomena such as processing speed cannot be measured offline: in a corpus design, there is no dependent variable which allows us to determine whether the hearer needs more time for the processing of constituent *y* in context *a* than in context *b*. And yet it is pivotal that we should find independent evidence for (14): it is only after we have established that some adjuncts and/or verbs facilitate the processing of the subject to a higher extent than others, that we can study *er*'s impact on subject processing in the context of adjuncts and verbs which do not intrinsically facilitate subject processing.

The self-paced reading (or subject-paced reading) technique we used in Grondelaers & Brysbaert 1996; and Grondelaers 2000: 197ff.) offers reliable online indications of processing difficulty or speed. Self-paced reading is a test paradigm in which participants read a sentence or short text on a computer screen, while pressing the spacebar to control the presentation of the successive segments of the sentence or text. In all our designs, the experimental stimuli were presented as in (16). The first screen contains a sequence of dots which corresponds to the letters of the words of the test sentence. A spacebar hit produces the first segment, which disappears when the participant taps the spacebar to indicate that he has read it; simultaneously, the next segment appears on the screen, which disappears with the next spacebar tap, until the participant has read the whole sentence. The time interval between two spacebar taps (i.e. the time it takes the participant to read the segment) is registered in milliseconds and recorded in a database, the underlying assumption being "that this measure reflects the time taken to execute at least some of the major processes associated with analysing the material in the display" (Mitchell 1984: 70).

(16)
```
-- --- ------- --- --- -------.
In het tuintje --- --- -------.
-- --- ------- was --- -------.
-- --- ------- --- een plataan.
```

If (14) is correct, then we can expect that *een zoutvat* "a salt tub" is read faster in (17) than in (18):

(17) In de keukenkast was een zoutvat
"In the kitchen cupboard was a salt tub."

(18) In het toneelstuk was een zoutvat
"In the play was a salt tub."

In de keukenkast "in the kitchen cupboard" referentially restricts the subject to the set of material objects associated with the preparation of food which are small enough to fit in a kitchen cupboard; since this category is not very large, and salt tubs represent a prototypical member of it, the subject in (17) will be processed effortlessly. In (18), by contrast, the adjunct confines the subject referent to the class of "objects, persons, and phenomena associated with the theatre", a category which is so large and heterogeneous that marginal stage props such as a salt tub will in all likelihood confound the hearer. This confusion, we assume in (14), will surface in a relatively longer processing time.

Grondelaers & Brysbaert (1996) report a self-paced reading experiment which confirms this expectation. We compared subject reading times in 24 sentences, which were divided in two sets on the basis of the variable "reference point potential of the adjunct": 12 sentences were constructed with a high-constraint adjunct such as *in het koffertje* "in the little suitcase", which code a three-dimensional, material container, 12 sentences with a low-constraint adjunct such as *in het toneelstuk* "in the play", which coded metaphorical inclusion in vague or abstract unbounded spaces; 54 undergraduate students from the Faculty of Arts of the University of Leuven – all native speakers with normal or corrected eyesight – participated voluntarily (for design details and experimental results, see Grondelaers & Brysbaert 1996).

The statistical analysis of the reading data revealed a facilitating impact on subject processing of high constraint adjuncts (*high constraint* 778 ms. < *low constraint* 832 ms., $p1$ & $p2 < 0.05$). This effect – which was replicated in three other experiments – clearly demonstrates that the accessibility of indefinite subject NP's is variable but – crucially – can be manipulated. Even relatively minor manipulations within an adjunct NP lead to perceptible subject processing differences: high-constraint adjuncts restrict the referential potential of the subject which, as a result, is processed faster. Low-constraint adjuncts, which do not, inhibit subject processing. Therefore, we can now proceed to the empirical substantiation of the hypothesis outlined in (15).

5. Corpus parameters of *er*'s inaccessibility marking function

In the previous paragraph we suggested that locative adjuncts are more efficient reference points than temporal adjuncts because they generate more useful predictive inferences about the forthcoming subject. The latter, however, specifically goes for locative adjuncts which denote concrete three-dimensional locations such as *keukenkast* "kitchen cupboard" or *ijskast* "refrigerator", which heavily constrain their subjects with respect to size and dimensionality. The majority of locative adjuncts, however, do not refer to such state-of-the-art reference points; in the light of the hypothesis in (15), it would be instructive to discover whether ADJUNCT CONCRETENESS – a potential operational definition of Langacker's *intrinsic salience* – correlates negatively with a preference for *er*. In the same vein, it would be revealing to find out whether Langacker's "contextually determined salience" equates ADJUNCT TOPICALITY (as suggested in for instance Taylor 1996: 209). Do adjunct topicality and *er*-preference also correlate negatively?

Notice that the latter is extremely plausible in our reference point account of adjunct sentences: a highly topical adjunct "entrenches" the adjunct sentence in the preceding context, as a result of which the subject constraining potential which emanates from the adjunct is extended to the preceding sentences, which considerably increases the subject reducing potential of the adjunct clause. Consider in this respect the following sentence:

(19) The visitors of the folk-museum feasted their eyes on the beautifully restored farmhouse; on the rough, oak-wood table stood a basket of fresh apples and *suspended from the roof-beam hung an enormous ham.*

Without the context, the constraining potential of the italicized adjunct clause is relatively restricted, since many things can be suspended from roof-beams. In the context of (19), however, *roof-beam* is metonymically inferred from *farmhouse* (and indirectly also from *folk-museum*), which dramatically reduces the class of objects which can be suspended from roof-beams. The first clause focuses the hearer's attention on the type of historical farmhouses he knows from period movies or novels; *a basket of apples* in the next clause zooms in on, or strongly suggests that the narrative will continue with, the gastronomic aspects of former farming. The adjunct *from the roof-beam* directs the comprehender's attention to the category of "foodstuffs associated with former farming which are suspended from roof-beams", and this category virtually coincides with the nominal type "ham".

In order to test both hypotheses, we extracted all locative adjunct sentences from a selection of Belgian and Netherlandic sources from the ConDiv-corpus (recall that the statistical significance of the factor REGION in the first regression forces us to carry out separate analyses of *er*'s distribution in Belgian and Netherlandic Dutch). We found 555 Belgian adjunct sentences in the Belgian UseNet logs, 1 Belgian nation-wide popular newspaper (*Het Laatste Nieuws*), and 1 Belgian nation-wide quality paper (*De Standaard*), as well as 297 Netherlandic adjunct sentences extracted from a corpus of Netherlandic UseNet logs, a Netherlandic nation-wide popular paper (*De Telegraaf*), and a Netherlandic nation-wide quality paper (*NRC Handelsblad*).

The factor ADJUNCT CONCRETENESS was implemented on the basis of the *salience taxonomy* developed in Grondelaers (2000: 176–177), an operational parameter which breaks up into 5 distinct values. The maximal value was assigned to adjuncts which refer to the physical aspects of the speech situation, by definition the most salient and stable reference point available to the language user. Value 2 was assigned to bounded two- or threedimensional concrete entities, and value 3 to unbounded concrete entities spaces. Value 4 was reserved for locations which cannot be straightforwardly categorized as either concrete or abstract. In a sentence such as *Bij de BOB is er een gebrek aan personeel* "At the Belgian Criminal Investigation Bureau there is a staff shortage", *de BOB* "The Belgian Criminal Investigation Bureau" refers to a metaphorical location which can nevertheless be metonymically interpreted as referring to the "the building in which the Bureau resides". Value 5, finally, was restricted to abstract locations which allow no spatial interpretation whatsoever.

Adjunct topicality was operationalized on the basis of a newly developed given-new taxonomy with ten values (space limitations preclude an in depth description, but details can be found in Grondelaers 2000: 156).

Table 4 contains Odds Ratios, p-values, and confidence intervals for four factors (register and verbal specificity were retained from the previous analysis):

Table 4. P-values, Odds Ratios and confidence intervals for 4 potential determinants of *er*'s distribution in Netherlandic (N) and Belgian (B) adjunct sentences

	N				B			
	p	**O.R.**	**conf. interv.**		**p**	**O.R.**	**conf. interv.**	
register	0.323	0.81	0.53	1.24	0.000	0.68	0.57	0.81
topicality	0.013	1.28	1.05	1.56	0.001	1.37	1.19	1.57
concreteness	0.052	1.48	1	2.21	0.002	1.45	1.09	1.94
specificity	0.000	0.17	0.1	0.3	0.003	0.23	0.16	0.32

Observe first that adjunct concreteness and adjunct topicality are not only significant determinants of *er*'s distribution in Belgian Dutch, but also in Netherlandic Dutch. Although in the latter, locative adjunct sentences are but rarely attested with *er*, it is typically abstract and discourse-new locations which trigger it. More significantly, in spite of the huge differences observed between the Belgian and Netherlandic distribution of *er* in the first regression, the extended Belgian and Netherlandic models appear to be *highly comparable* as far as the relative impact of the different variables is concerned. The regression returns *confidence intervals* for the Odds Ratios (delimiting the interval in which the Odds Ratio finds itself, given the variance in the data and the significance level) which can be used to assess the degree of overlap between two models with identical ingredients. When we compare the Netherlandic and Belgian confidence intervals, it turns out that they overlap for each variable, which gives us statistical evidence that there are no proportional differences between the Belgian and the Netherlandic models.

However, when we focus on the respective success rates of the Netherlandic and Belgian models, we find that although the addition of both adjunct factors noticeably improves their predictive power, the effect is *much more outspoken* in the Belgian model: whereas the Netherlandic model gains 6.78% of predictive power (going from 79.32% – without factors – to 86.1%, with the factors we added), the Belgian model proceeds as much as 13.61%, to end at a success rate comparable to that of the Netherlandic model (from 70.96% to 84.57%)!

So all in all, we now have good reason to assume that *er* functions as an inaccessibility marker in Netherlandic as well as in Belgian Dutch. The differences between both distributions – *er* is used much less in Netherlandic than in Belgian Dutch, and its distribution obeys much more clear-cut rules – can be attributed, we believe, to the different linguistic standardization of Belgian and Netherlandic Dutch: whereas Netherlandic Dutch benefited from a normal standardization, the standardization of Belgian Dutch was blocked in the 16th century as a result of political and social factors, resuming its course only in the 20th century (see Geeraerts, Grondelaers, & Speelman 1999: Chapter 2) for a concise historical overview). Observe to begin with that the register differences which characterize *er*'s distribution in Belgian but not in Netherlandic Dutch are characteristic of a delayed standardization (see in particular Geeraerts, Grondelaers, & Speelman 1999; and Grondelaers, Van Aken, Speelman, & Geeraerts 2001, which contain corpus-based evidence to the effect that register differences in Belgian Dutch are much larger than in Netherlandic Dutch). On a more speculative note, it also seems that the Netherlandic use of *er* represents an evolved, "grammaticalized" stage of the Belgian *er*-preferences. In Netherlandic Dutch, the use of *er* – which is determined by more formal distinctions with clearer cut-off points – appears to have been "wired" to some extent into the grammatical competence of the language user (the term "competence" is used in a strictly non-technical sense here). In Belgian Dutch, by contrast, the distribution of *er* suggests a non-conventionalized, highly individual selection process which proceeds on a case-per-case basis, engendering lots of indecision and variance.

6. Conclusion

The adoption of the corpus methodology illustrated here has a number of consequences for the organization of linguistic research that may serve as a conclusion to this chapter.

First, linguistic research is likely to become more *collaborative*. Acquiring and maintaining corpora, developing tools, exploring analytical techniques presupposes teamwork rather than an individual approach. Progress in this area is likely to come not from purely individual efforts, but from specialized local groups and highly interactive networks.

Second, linguistic research will become much more *cumulative*. When linguistic hypotheses can be tested against a shared basis of corpus data, they will become more comparable than is currently the case, with many theories existing in parallel, without sufficient common ground for a stringent comparison of competing models.

And third, linguistic research will become *slower*. The method of gradually refining interpretative hypotheses through a recurring confrontation with empirical data is a painstaking one, and the study of language will have to suppress its tendency – all too conspicuous in modern linguistics – to jump to grand but sparsely substantiated theories.

References

Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Barlow, M. & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford: CSLI.

Bays, H. (1998). Framing and face in internet exchanges: A socio-cognitive approach. Online publication on http://viadrina.euv-frankfurt-o.de/~wjournal/bays.html.

Bolinger, D. (1977). *Meaning and form*. London: Longman.

Bybee, J. & Hopper, P. (Eds.). (2001a). *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Bybee, J. & Hopper, P. (2001b). Introduction. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 1–24). Amsterdam: John Benjamins.

Croft, W. & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Dirven, R. & Taylor, J. (1988). The conceptualisation of vertical space in English: The case of *Tall*. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics* (pp. 379–402). Amsterdam: John Benjamins.

Es, G. A. van & Caspel, P. P. J. van (1971). De patronen van de zinspotente groepen; Grondtype A en zijn varianten II. *Publicaties van het archief voor de Nederlandse syntaxis*. Groningen: Rijksuniversiteit.

Geeraerts, D. (1985). *Paradigm and paradox. Explorations into a paradigmatic theory of meaning and its epistemological background*. Leuven: Universitaire Pers.

Geeraerts, D. (1999). Noch standaard, noch dialect. 'Tussentaal' in Vlaanderen en Nederland. *Onze Taal, 68,* 232–235.

Geeraerts, D., Grondelaers, S., & Bakema, P. (1994). *The structure of lexical variation. Meaning, naming, and context.* Berlin: Mouton de Gruyter.

Geeraerts, D., Grondelaers, S., & Speelman, D. (1999). *Convergentie en divergentie in de Nederlandse woordenschat. Een onderzoek naar kleding- en voetbalnamen*. Amsterdam: Meertensinstituut.

Goossens, L. (1990). Metaphtonymy: The interaction of metaphor and metonymy in expressions for linguistic actions. *Cognitive Linguistics, 1,* 323–340.

Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics: A study of particle placement.* London: Continuum Press.

Grondelaers, S. (2000). *De distributie van niet-anaforisch* er *buiten de eerste zinsplaats. Sociolectische, functionele en psycholinguistische aspecten van* er's *status als presentatief signaal.* PhD Thesis, University of Leuven.

Grondelaers, S. & Brysbaert, M. (1996). De distributie van het presentatieve *er* buiten de eerste zinsplaats. *Nederlandse Taalkunde, 1,* 280–305.

Grondelaers, S., Deygers, K., Van Aken, H., Van den Heede, V., & Speelman, D. (2000). Het CONDIV-corpus geschreven Nederlands. *Nederlandse Taalkunde, 5,* 356–363.

Grondelaers, S., Speelman, D., & Geeraerts, D. (2002). Regressing on *er*. Statistical analysis of texts and language variation. In A. Morin & P. Sébillot (Eds.), *6th International Conference on the Statistical Analysis of Textual Data* (335–346). Rennes: Institut National de Recherche en Informatique et en Automatique.

Grondelaers, S., van Aken, H., Speelman, D., & Geeraerts, D. (2001). Inhoudswoorden en preposities als standaardiseringsindicatoren. De diachrone en synchrone status van het Belgische Nederlands. *Nederlandse Taalkunde, 6,* 179–202.

Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. C. (1997). *Algemene Nederlandse Spraakkunst*. Groningen & Deurne: Martinus Nijhoff – Wolters Plantyn.

Hüllen, W. & Schulze, R. (Eds.). (1988). *Understanding the lexicon: Meaning, sense and world knowledge in lexical semantics.* Tübingen: Max Niemeyer Verlag.

Istendael, G. Van (1993[10]). *Het Belgisch labyrinth. Wakker worden in een ander Land*. Amsterdam: De Arbeiderspers.

Kemmer, S. & Barlow, M. (2000). Introduction. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (vii–xxviii). Stanford: CSLI.

Kirsner, R. S. (1979). *The problem of presentative sentences in Modern Dutch*. Amsterdam: North Holland.

Lakoff, G. (1987). *Women, fire, and dangerous Things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Langacker, R. W. (1990). *Concept, image and symbol. The cognitive basis of grammar.* New York: Mouton de Gruyter.

Langacker, R. W. (1991). *Foundations of Cognitive Grammar. Volume 2: Descriptive application.* Stanford: Stanford University Press.

Langacker, R. W. (1993). Reference point constructions. *Cognitive Linguistics, 4,* 1–38.

Levin, B. (1993). *English verb classes and alternations. A preliminary investigation*. Chicago & London: The Chicago University Press.

Mitchell, D. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In D. E. Kieras & M. Just (Eds.), *New methods in reading comprehension research* (pp. 69–89). Hillsdale: Erlbaum.

Morin, A. & Sébillot, P. (Eds.). (2002). *6th International Conference on the Statistical Analysis of Textual Data*. Rennes: Institut National de Recherche en Informatique et en Automatique.

Rietveld, T. & Van Hout, R. (1993). *Statistical techniques for the study of language and language behaviour*. Berlin: Walter de Gruyter.

Rohdenburg, G. & Mondorf, B. (Eds.). (2003). *Determinants of grammatical variation*. Berlin: Mouton de Gruyter.

Rooij, J. De (1991). Regionale variatie in het gebruik van er III. *Taal en Tongval, 43,* 113–136.

Rudzka-Ostyn, B. (Ed.). (1988). *Topics in cognitive linguistics*. Amsterdam: Benjamins.

Schulze, R. (1988). A short story of *down*. In W. Hüllen & R. Schulze (Eds.), *Understanding the lexicon: Meaning, sense and world knowledge in lexical semantics* (pp. 394–410). Tübingen: Max Niemeyer Verlag.

Schwanenflugel, P. J. (1986). Completion norms for final words of sentences using a multiple production measure. *Behavior Research Methods, Instruments, & Computers*, *18,* 363–371.

Speelman, D., Grondelaers, S., & Geeraerts, D. (2003). Profile-based linguistic uniformity as a generic method for comparing language varieties. *Computers and the Humanities, 37,* 317–337.

Stefanowitsch, A. (2003). Constructional semantics as a limit to grammatical alternation: The two genitives of English. In G. Rohdenburg & B. Mondorf (Eds.), *Determinants of grammatical variation* (pp. 413–444). Berlin: Mouton de Gruyter.

Stefanowitsch, A. & Gries, S. T. (2003). Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics, 8,* 209–243.

Taeldeman, J. (1992). Welk Nederlands voor de Vlamingen? *Nederlands van Nu, 40,* 33–52.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 49–100.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics, 11,* 61–82.

Taylor, J. (1996). *Possessives in English. An exploration in Cognitive Grammar*. Oxford: Clarendon.

Verhagen, A. & van de Weijer J. (Eds.). (2003). *Usage-based approaches to Dutch*. Utrecht: LOT.

Verkuyl, H. J. (1998). O corpora, O mores. *Nederlandse Taalkunde, 3,* 60–63.

PART III

Sign language and gesture

Empirical methods in signed language research

Sherman Wilcox and Jill P. Morford

1. Background on signed languages and cognitive linguistics

1.1 Linguistic research on signed languages

Signed languages are natural human languages used by deaf people throughout the world as their native or primary language.[1] The 13th edition of the Summer Institute of Linguistics *Ethnologue* of the world's languages lists 114 signed languages (Grimes 1996), although linguists generally assume that they number well into the hundreds. The gestural-visual modality of signed languages is reflected in their linguistic structure. Signed languages make extensive use of space, for example by incorporating spatial locations to indicate verbal arguments. In addition to the hands, the face plays a critical role in signed language grammar, expressing a range of information such as questions, topic, adverbials, and so forth.

Over the past 40 years, linguists have demonstrated that signed languages such as American Sign Language (ASL) may be analyzed and described using the same units as spoken languages. While differences in structure attributable to modality (spoken versus signed) have been noted, the overwhelming conclusion is that signed languages share important characteristics with spoken languages.

The modern era of linguistic research on signed language began in the late 1960s with the pioneering work of the American linguist William C. Stokoe. Stokoe was a professor of English at Gallaudet College (now Gallaudet University) in Washington DC, the only liberal arts university for deaf people. Stokoe began to apply linguistic techniques borrowed from the structuralist tradition prevalent at the time to study the language that he saw deaf students using in his classroom – ASL. His research eventually led to a broad interest in the structure of ASL by linguists and initiated research worldwide to analyze the world's signed languages.

A common misunderstanding is that signed languages are merely representations of spoken languages – that ASL, for example, is a signed representation of spoken English. Signed languages are independent languages with their own lexicons and grammars. Like

1. Many authors refer to 'sign languages'. We prefer the term signed language, parallel to spoken language and written language.

spoken languages, signed languages have genetic and historical relations with other signed languages. ASL's closest genetic relative, for example, is French Sign Language (LSF).

1.1.1 *Phonology*

One of the pioneering discoveries made by Stokoe was that ASL can be described phonologically. Before this, it was assumed that the signs – that is, the words – of a signed language were unanalyzable. Stokoe showed that a sign consists of analyzable units of structure and coined the term *chereme* for these structural units (Stokoe 1960).

Stokoe analyzed the phonology of signs into three major classes: handshape (the configuration that the hand makes when producing the sign), location (the place where the sign is produced, for example on the head, or in the neutral space in front of the signer's body), and movement (the movement made by the signer's hands in producing the sign, for example upward or towards the signer's body). Stokoe called these *aspects* of a sign. Later linguists called these aspects the parameters of a sign and added a fourth parameter: orientation, the direction which the palm of the hand faces when producing the sign (Battison 1978). One form of evidence for these parameters is demonstrated by the existence of minimal pairs, signs differing only in one parameter which have different meanings (Klima & Bellugi 1979).

1.1.2 *Morphology*

ASL, like many signed languages, is highly synthetic with tendencies towards polysynthesis. ASL allows morphemes indicating action, person agreement, aspect, and adverbial information to be combined into a single, multimorphemic ASL word; for example, 'I very carefully gave [one] to each [person]' would be expressed with a single sign in ASL.

1.1.3 *Syntax*

Research on the syntax of signed languages has examined issues of word class, word order, and relations among constituents such as relative clauses, question formation, topic-comment structure and the flow of information in discourse, and the grammatical use of space.

1.1.4 *Fingerspelling*

Another common misunderstanding is that signed languages are merely (or largely) comprised of fingerspellings. This is not the case. Fingerspelling makes use of handshape configurations that correspond to the alphabet of the majority written language. Fingerspelling is often used for proper names or technical terms, and is used for loan words in signed languages; for example, *of, all, sure*, and several other English words have been borrowed into ASL through fingerspelling. A variety of fingerspelling systems exist among the world's signed languages. ASL and many other signed languages use a one-handed system. British Sign Language (BSL) uses a two-handed fingerspelling system. The amount of fingerspelling used in a signed language varies greatly. ASL and BSL rely extensively on fingerspelling; the use of fingerspelling in most other signed languages is more restricted.

Fingerspelling is more than a sequence of canonical handshape configurations, since the articulatory movements within the fingerspelled word influence each other. Perseverative and anticipatory coarticulation affects the actual shaping of fingerspelled words, creating a fluid transition between letters (Wilcox 1992).

1.2 Cognitive linguistics

The cognitive linguistic approach to language makes assumptions that are quite different from structuralist or generative approaches that drove so much of the early work on signed languages. Because of this, cognitive linguistics permits the linguist to examine questions relevant to signed languages that were left unexplored under previous theories. In turn, signed languages, because of their visual modality, permit linguists to address central issues in cognitive linguistics in new and unique ways.

Croft and Cruse (2004: 1) identify three major hypotheses that guide the cognitive linguistic approach to language:

- Language is not an autonomous cognitive faculty
- Grammar is conceptualization
- Knowledge of language emerges from language use

The first hypothesis suggests that knowledge of language is no different than knowledge in general. In studying language, linguists working within a cognitive linguistic approach necessarily are interested in all types of knowledge, including models of memory, perception, and attention; the organization of knowledge into categories, prototypes, frames, schemas; the organization of conceptual structure into domains and spaces; and the dynamic cognitive processing that occurs whenever humans use language, such as coding and construal operations including selection, point of view, and figure-ground organization. Cognitive linguistics claims that these knowledge structures are based on embodied archetypes resulting from our lived interactions with the world, such as:

- Conceptions of objects moving in three-dimensional space
- Conceptions of the energetic interactions of objects and the transfer of energy
- Conceptions of viewing scenes from certain perspectives or vantage points
- Conceptions of real-world effects on our bodies, such as force dynamics and barriers

One striking feature of the cognitive linguistic framework is the remarkable degree to which visual perception enters into accounts of grammar. This is one area where signed languages can offer new insight, by providing data in the visual domain complementary to spoken language data.

The second hypothesis recognizes that it is not just within the realm of semantics that a cognitive approach to language is relevant: all grammar is seen as essentially symbolic. The lexicon, morphology, and syntax form a continuum of symbolic elements that provide for the linguistic structuring and construal of conceptual content.

The third hypothesis makes the critical assumption that the categories and conceptual structures that comprise grammar are derived from our conception of specific utterances, actual uses. In order to account for the grammar of a language, the cognitive linguistic approach demands that we understand how general cognitive processes of schematization, abstraction, and generalization work to build up a conceptual model of language. As Croft and Cruse (2004: 4) note: "cognitive linguists argue that the detailed analysis of subtle variations in syntactic behavior and semantic interpretation give rise to a different model of grammatical representation that accommodates idiosyncrasies as well as highly general patterns of linguistic behavior."

These three foundational hypotheses of cognitive linguistics have special relevance for signed language linguists, and conversely, the study of signed languages plays a critical role in the development of cognitive linguistic theory. Cognitive linguistics is important for signed language linguists because it addresses language in a radically different way than other linguistic theories, affording linguists the opportunity to explore the linguistic and conceptual structures that are clearly relevant for signed languages yet have no theoretical status in other linguistic theories.

Conversely, cognitive linguistics challenges signed language researchers in a number of important ways. For example, under previous approaches, it was possible to conduct studies of phonology and syntax independently of any extensive knowledge of meaning. Because of the central importance of meaning and conceptualization, a cognitive linguistic approach demands a detailed knowledge of lexical relations, polysemy, hyponymy, antonymy, metaphor, and metonymy in order to conduct grammatical studies of signed languages.

Cognitive linguistics also requires the researcher to understand the details of actual language use and frequency. This poses exceptional data collection and analysis challenges for the signed language linguist which we will discuss in Section 3 below.

2. Signed language research in a cognitive linguistics framework

Because of its focus on conceptual structures, cognitive linguistics drives research in such areas as metaphor, metonymy, and iconicity and how they motivate lexical, morphological, and syntactic behavior, as well as their role in the process of grammaticization. Gesture studies have also recently become an area of great interest for cognitive linguists because of the rather direct way that gestures can provide insight into cognitive structures and conceptualizations (cf. Sweetser this volume). As we will discuss, a rich area of empirical study lies in the intersection of cognitive linguistic approaches to gesture and signed languages.

Some of the earliest work on signed languages investigated precisely these issues (see Friedman 1977). These studies were provocative in that they suggested that iconicity, analog coding, metaphor, and metonymy played a crucial role in the grammar of ASL. However, they were also largely ignored by the field of signed language linguistics, which was pre-occupied at the time with demonstrating the parallels between signed and spoken languages in an effort to achieve legitimacy in the broader field. This situation has changed with the emergence of cognitive linguistics, and the wide acceptance of signed languages as linguistic systems. Linguists are once again able to turn their attention to these areas of research. A body of work now exists on iconicity, metaphor, mental spaces, gesture and grammaticization, and language evolution. A few examples from each area will illustrate the unique perspective that the study of signed languages brings to cognitive linguistics.

2.1 Iconicity

Signed languages, because of their visual modality, would seem to be particularly productive sources of data for the study of iconicity. Indeed, Stokoe et al. (1965) noted metaphorical, metonymic, and iconic aspects of signed languages. Iconicity, in particular, was seen to be a prominent feature of signed languages.

Much of the early work on ASL, while it recognized the pervasive presence of iconicity, devoted its attention more to reducing the significance of iconicity than to studying what it tells us about grammatical structure. For example, studies examined the loss of iconicity over time in the lexicon (Frishberg 1975), demonstrated that iconicity does not appear to play a role in language acquisition (Meier 1980), and concluded that iconicity does not aid in the processing of signs (Klima & Bellugi 1979).

Recent research is calling into question some of these claims about the role of iconicity in signed languages. For example, Grote and Linz (2003) have addressed whether the form of signs influences a signer's representation of the semantic domain of a sign's referent, particularly in the case of iconic signs. They hypothesize that the iconic relationship between a sign and its referent is not created purely via a likeness, but is the result of "a mentally constructed correspondence between two cognitive products" (25). They describe two experimental tasks investigating this issue. In the first, deaf signers of German Sign Language (Deutsche Gebärdensprache – DGS), hearing speakers of spoken German, and hearing bilinguals (German/DGS) were presented with a DGS sign or a German word, followed by a picture, and were asked to decide whether there was a semantic relationship between the two. Some of the pictures represented the iconic base of a DGS sign, for example, the beak of an eagle. The sign EAGLE in DGS is made by showing the shape of the beak with the index finger at the nose. Other pictures represented semantic features unrelated to the iconic base of a sign, for example, the claws or the wings of an eagle. Finally, some pictures were completely unrelated to the sign or word they were paired with (see Figure 1). Participants responded yes or no to indicate whether they thought the picture was related to the preceding word or sign. Evaluating only "yes" responses, Grote and Linz found that the deaf and hearing signers were faster to respond to pictures of the iconic base of a sign than to pictures that depicted other semantic features of the sign's meaning. Interestingly, the same pattern of results was found when the bilinguals were tested in German. For example, bilinguals were faster to conclude that a picture of a beak was related to the German word *Adler* (eagle), than they were able to respond when a picture of a wing followed the same word *Adler*. No differences in response time were found for the hearing participants with no knowledge of DGS.

In a second task, the same participant groups were presented with a word or sign followed by two pictures. In this case, participants were asked to decide which of the two pictures was more closely semantically related to the word or sign. For example, participants heard the word *Kuh* (cow) or saw the DGS sign COW, which is made by tracing the form of a horn at the forehead. Subsequently, participants saw a picture of a cow's head with horns, or a cow's body with a black and white pattern of coloring. Results replicated the findings in the first task. If the picture depicted the iconic base of a sign (the horns of a cow) rather than another semantic element (cow's body), responses were faster for

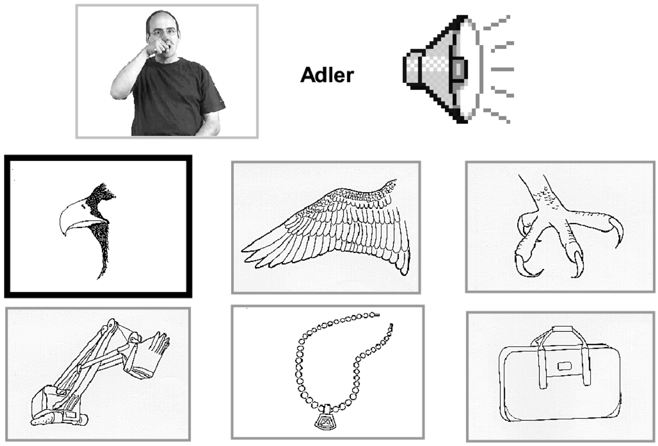


Figure 1. Pictures presented with the DGS sign for 'eagle' and the spoken German word 'Adler'

deaf and hearing signers, but not for hearing speakers of German. Again the bilinguals showed the same pattern of responses even when performing the task in German. These studies both indicate that the iconic base of a sign is prominent in the conceptualization of the sign's referent. Because the results are identical when bilinguals are tested in their second language, the authors conclude that sign-mediated effects on conceptualization are not language-specific (but see Iversen, Morford, Nuerk & Willmes 2004; and Iversen, Nuerk & Willmes 2004 for evidence of language-specific sign-mediated effects on number processing).

Despite newer studies of this nature, the role of iconicity in signed language grammar is still quite controversial. An indication of how far the rejection of iconicity in grammar has gone is exemplified in Valli and Lucas (1995: 7):

> It is probably true that the form of the sign SIT is an iconic representation of human legs sitting ...[However,] focusing on its iconicity will not provide much insight into the interesting relationship between SIT and the noun CHAIR, and other noun-verb pairs. Nor will [iconicity] help explain how the movement of SIT can be modified to mean SIT FOR A LONG TIME (slow, circular movement) or SIT ABRUPTLY (short, sharp movement).

Wilcox (2002) demonstrates that in fact the grammatical categories of noun and verb in ASL noun-verb pairs do exhibit an iconic relation (cf. Johnston 2001). Nouns profile a bounded region in space both semantically and phonologically. Verbs profile phonological movement in space; semantically, movement in space is a prototypical example of a process, a defining property of the category verb (Langacker 1987). The descriptions of the two aspectual verb forms are striking in how transparently they demonstrate the iconic relation between meaning and form: slow, circular movements indicating the durative form and short, sharp movements indicating the punctual form.

Even those linguists who did acknowledge iconicity generally accepted that it played little role in the grammar of ASL. Klima and Bellugi (1979: 70), for example, while acknowledging the "two faces of signs: the iconic face and the encoded, arbitrary face" concluded that (ibid.: 79):

> The iconic face does not show at all in the processing of signs in immediate memory. Historical change diminishes the iconic properties of ASL signs; some signs become more opaque over time, some completely arbitrary.

Further, Klima and Bellugi claimed that the grammar of ASL works to override or submerge the inherent iconicity of signs:

> Grammatical operations that signs undergo can further submerge iconicity. Thus many signs, while having their roots deeply embedded in mimetic representation, have lost their original transparency as they have been constrained more tightly by the linguistic system. (ibid.: 34)

The example that Klima and Bellugi offered is the morphological marking of intensification on certain statives in ASL, expressed phonologically as an initial hold of the sign's movement followed by sudden, rapid release. When this grammatical marker appears on the ASL sign SLOW, the resulting sign means 'very slow'. Klima and Bellugi pointed out that the sign VERY-SLOW is made with a fast movement – faster than that used in the sign SLOW: "Thus the form of 'very slow' is incongruent with the meaning of the basic sign" (1979: 30). It is this fact that supported their claim that the grammar has submerged iconicity.

Using a cognitive iconicity analysis, Wilcox (2004a) comes to a different conclusion about whether grammar submerges iconicity. Cognitive iconicity is based on the theoretical model of cognitive grammar (Langacker 1987). A critical claim of cognitive grammar is that both semantic and phonological structures reside within a language user's conceptual space. Conceptual space is multidimensional, encompassing all of our thought and knowledge, "the multifaceted field of conceptual potential within which thought and conceptualization unfold" (Langacker 1987: 76). By adopting this view we can talk about similarities as distance between structures that reside in multi-dimensional conceptual space (Gärdenfors 2000). Certain notions reside close to each other in conceptual space because they possess certain similarities. Other notions reside farther apart in conceptual space, reflecting their dissimilarity.

What is important for understanding cognitive iconicity is the claim that phonological notions also reside in conceptual space. The phonological pole of symbolic linguistic structures reflects our conceptualization of pronunciations, which range from the specific pronunciation of actual words in all their contextual richness to more schematic conceptions, such as a common phonological shape shared by certain nouns or verbs, such as those participating in noun-verb pairs discussed above, in a particular language.

The typical case for language is that the semantic and the phonological poles of a symbolic structure reside in vastly distant regions of conceptual space. The sound of the spoken word *dog,* for example, has little in common with the meaning of the word. This great distance in conceptual space, and the resulting incommensurability of the semantic and phonological poles, is the basis for Saussure's contention of *l'arbitraire du signe*. Alternatively, when the phonological and semantic poles of signs reside in the same region of conceptual space, arbitrariness is reduced.

Cognitive iconicity thus is defined not as a relation between the form of a sign and real world referent, but as a distance relation within our multidimensional conceptual space between the phonological and semantic poles of symbolic structures. Studying iconicity under this view thus requires the researcher not only to understand how the semantic pole of a linguistic unit is conceptualized, but also how the phonological pole is conceptualized. Limiting our discussion only to the hands, certain conceptual properties of signed language articulators are discernable:

– The hands are autonomous objects manifest in the spatial domain.
– Movement is a dependent property of handshapes, manifest in the temporal domain.
– Location is a dependent property, manifest in the spatial and temporal domain.
– Orientation is a dependent property of handshapes, manifest in the spatial domain.

These properties of signed language articulators lead naturally to conceptualizations such as:

– Hands may be conceptualized as objects with shapes
– Hands may be conceptualized as objects that move in space
– Hands may be conceptualized as objects performing some function

Returning to the question of whether grammatical processes in ASL submerge iconicity, Wilcox (2004a) notes that VERY-SLOW is multimorphemic, consisting of a root lexical morpheme SLOW and a bound, grammatical morpheme marking intensification. The same bound morpheme appears on other lexical roots, such as VERY-SMART and VERY-FAST.

Thus, while it is true that the form of VERY-SLOW is incongruent with the meaning of the lexical root SLOW, it is not the case that the form of the intensifier morpheme is incongruent with its meaning. In fact, it is iconic. To see this, Wilcox notes two facts about the conceptualization of intensity. Intensity is a conceptually-dependent notion: intensity depends on a prior conception of what is being intensified. Something is 'very *slow*' or 'very *big*' but not simply 'very'. Second, the abstract notion of intensity is often understood metaphorically by reference to more grounded concepts such as the build-up and sudden release of internal pressure, as happens when we shake a soda can and then open it.[2]

How then is the form VERY-SLOW iconic? First, it is iconic because the articulators directly represent the metaphoric conceptualization of intensity as a sudden release of pent-up pressure: the phonetic realization of this bound morpheme is an initial hold followed by the sudden release of the lexical morpheme's movement. Second, intensity as a conceptually-dependent notion is iconically represented: change in *how* the sign's move-

2. Kövecses (2000) notes that one folk understanding of anger involves a cognitive model in which intensity of offense outweighs intensity of retribution creating an imbalance that causes anger. As a result, a common cross-linguistic metaphorical expression of anger involves the conceptual metaphor ANGRY PERSON IS A PRESSURIZED CONTAINER.

ment is articulated is conceptually-dependent because it relies on a prior conception of *what* movement was produced in this way.

Work on iconicity in signed languages has been rejuvenated by the cognitive linguistic perspective. In addition to studying iconicity as an instance of conceptual mapping, researchers have begun to examine its relationship to other cognitive processes such as metaphor and metonymy (Brennan 1990; Taub 2001; P. Wilcox 2000; Wilcox, Wilcox & Jarque 2004) as well as variation in iconicity by discourse context (Russo 2004).

2.2 Metaphor

Although Stokoe pointed out early on that ASL incorporates metaphor, little was known about metaphor in ASL prior to the influence of cognitive linguistics. One problem was that metaphor was often confused with iconicity. It is easy to see why this might have occurred. If metaphor is "understanding and experiencing one kind of thing in terms of another" (Lakoff & Johnson 1980: 5), then representing long, thin, upright objects (e.g., blades of grass) by another set of long, thin, upright objects (the outstretched fingers of both hands, palms up) could be mistaken for metaphor.

One of the first to apply insights from the cognitive linguistic study of metaphor to ASL was Ronnie Wilbur (1987). Following the Lakoff and Johnson model, she identified a small set of metaphorical signs. For example, she noted that a HAPPY IS UP spatialization metaphor occurs in ASL signs such as CHEERFUL and HAPPY, which have an upward movement. The NEGATIVE IS DOWN metaphor occurs in LOUSY and FAIL, which have a downward movement.

The work of Phyllis Wilcox (2000) represents the first in-depth analysis of metaphor in ASL inspired by cognitive linguistics. In addition to clarifying the distinction between iconicity and metaphor, and between metaphor and other tropes such as simile and metonymy, Wilcox identified a number of structural and ontological metaphors in ASL. Even when these metaphors occurred as lexical items, they involved complex networks of metaphorical mappings. For example, Wilcox (2000: 109–124) demonstrated a set of ontological metaphors based on IDEAS ARE OBJECTS. This metaphor invites the entailment that ideas are subject to physical force, such as gravity. This occurs in one example from Wilcox's corpus in which a deaf person is describing memorizing the lines of a play. Something happens to distract him, and he describes the lines as 'falling out of his head'.

Another entailment is that ideas can be handled in various ways, grasped, and selected or discriminated. Wilcox demonstrates that each of these metaphorical concepts motivates a different use of an ASL classifier for handling objects. When ideas are conceptualized as objects that can be manipulated, ASL expressions use a flat-O morpheme, which is also used to represent manipulation of "thin, flattish, wide objects" such as a sheet of paper or a thin pamphlet. For example, using this sign, ideas can be placed at various locations in the head for later retrieval. A conventional ASL compound meaning "I learned my lesson" uses this sign made on the forehead, followed by an open hand, fingers together and touching the forehead, meaning 'to retain' or 'to stay put'.

Other forms of handle classifier signs are used for different conceptualizations of ideas. The S handshape, which is used to express holding onto an object for the purpose of

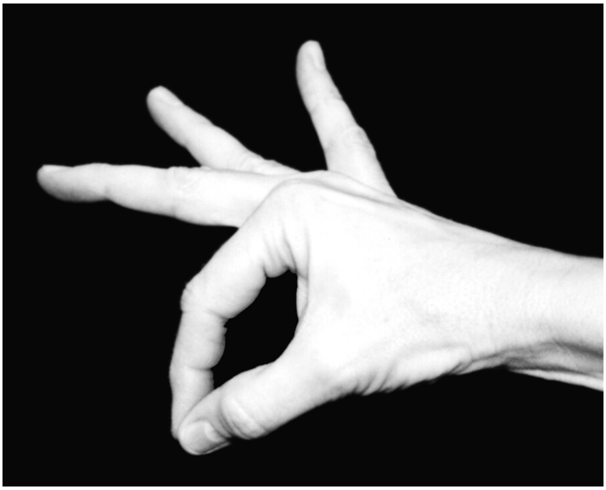


Figure 2. F-classifier handshape used to represent carefully selecting and handling a small object

retaining control, is seen when the ideas are to be retained: when discussing an idea, this sign can be made close to the forehead to mean "I'll keep that in mind." A related morpheme, expressing loss of control, is made by releasing the S handshape into an open-5 handshape. It is used, for example, when a physical object is held and then dropped. One of Wilcox's language consultants told her that he had traveled around the world collecting folklore. He had collected all of these stories in his head, with the intention of one day producing a historical document about Deaf folklore. To express this, he used the sign meaning 'to grasp and then release from control' a physical object. The initial portion of the sign (the S handshape, indicating holding and controlling) was made near the forehead; the final portion (the open-5 handshape, indicating release from his possession in the form of a written document) moved from the forehead location to a neutral location in front of the signer, where a book might be located.

Finally, Wilcox shows that when signers wish to express that ideas and concepts need to be carefully selected, they use a handle sign that means 'to carefully discriminate a small object from among many possible alternatives'. For example, a signer might use this sign, an F classifier handshape, to represent carefully selecting and handling a small object such as a small button or a small pin (see Figure 2). When discussing ideas, a signer might use this sign to express that she selects one particular idea from many that she has formed.

In one complex example, Wilcox reports that a consultant explained that it is important to place ideas in front of the body where they can be clearly seen. Then, rather than blindly selecting an idea, she can carefully select the appropriate idea. This was expressed with a series of signs:

1. CONTAINER_1-AT-FOREHEAD
2. CONTAINER_1-MOVES-TO-POSITION_2 [position in front of body]
3. Right hand: LOOK-AT_2 [look in the direction of the container Left hand: CONTAINER_1

4. Right hand: CAREFULLY-SELECT-FROM Left hand: $CONTAINER_1$

Another researcher who brings the cognitive linguistic framework to the study of signed language is Taub (2001). According to Taub, what is unique about ASL, and by implication about signed languages more generally, is the way that iconicity and metaphor combine to create a "double mapping" (96ff.). In spoken languages, a metaphorical expression maps language about a concrete domain onto an abstract domain. However, in a signed language, the iconicity of the signs drawn from the source domain can be directly manipulated such that metaphorical expressions are a combination of an iconic mapping between the source domain and the articulators, and a metaphorical mapping between the source domain and the target domain. Theoretically then, by manipulating the iconicity of the signs drawn from the source domain, novel utterances can be generated, uniquely suited to the signer's conceptualization. Taub claims that this is the case for ASL, which rarely uses the type of metaphorical expressions found in English, such as "I couldn't get my point across" (94), preferring instead to modify a sign like THINK, produced by moving the index finger in a small circle at the forehead, to generate signs such as "THINK-BOUNCE", which depicts an idea (in the form of the index finger), moving from the forehead, and then deflecting off a flat surface.

Also examining the relation between metaphor and iconicity, Bouvet (1996) investigates the semiotic basis of iconic signs in French Sign Language and argues that in many cases, the relationship between a sign and its referent is grounded in metaphorical projections of embodied representations. For concrete referents, such as food or household objects – referents with which humans have considerable perceptual and motor experience – this conclusion is perhaps predictable since the hands are used to represent objects that are frequently handled. But Bouvet demonstrates with particular care how signs for abstract concepts reveal a metaphorical projection of the conceptual domain onto the domain of the human body. These signs often have parallels in spoken language, but speakers tend not to be aware of them. Individuals who have had no previous contact with a signed language often assume that it is difficult or impossible to express abstract concepts in a signed language. On the contrary, Bouvet argues, signed languages are ideal for the expression of abstract concepts – and can draw our attention to the ubiquity of metaphorical projections in spoken languages as well.

As the study of metaphor in signed languages progresses, researchers are turning their attention from lexical to narrative manifestations of metaphor. P. Wilcox (2004) has demonstrated how complex metaphors, metonymies, and blends create coherence and cohesion in narrative discourse. Although others are also extending their analysis of metaphor, iconicity, and metonymy into the discourse realm, including researchers studying other signed languages (Brennan 2005; Russo et al. 2001; Russo 2005), clearly much work remains to be done to bring our knowledge up to the same level as exists for spoken languages.

2.3 Mental spaces and blends

Liddell (1995, 2000) identifies three mental spaces used in ASL discourse for establishing and maintaining references: real space, surrogate space, and token space. Real space is defined as the signer's conception of her current, directly perceivable physical environment. It is used to refer to objects directly perceivable in the signer's physical space, such as an interlocutor. One example is the ASL pronoun PRO, produced by directing the tip of a 1 handshape toward the referent "regardless of where the referent happens to be" (24).

Surrogate space is similar to real space, the difference being that surrogate space is invoked when signers make reference to people and things not physically present. Thus, the same sign PRO could be produced in surrogate space, but the referent is located relative to other signs in surrogate space rather than pointing to the actual location of the referent. Token space also consists of a conceptualization of a non-present entity. Token space differs from surrogate space in that surrogate space can use the physical space surrounding the signer, but token space is limited to the size of the physical space in front of the signer in which the hands are located while signing (Liddell 1995: 33). Entities in surrogate space are thus conceptualized and treated as normal sized and with body features. When a signer discusses a non-present person, surrogate space allows her to indicate areas on that person, such as the head or the feet. Entities in token space, however, are not normal sized and are treated as featureless. Liddell goes on to describe the nature of linguistic representations and grammatical phenomena characteristic of these three mental spaces. For example, he notes that surrogates are used to represent 1st, 2nd, and 3rd person roles in discourse, while tokens are limited to 3rd person only.

On the basis of his analysis Liddell presents a provocative claim about the relation between signed languages and gesture. Because the set of possible locations in these forms is open, Liddell argues that the handshape, orientation, and movement of these forms are lexically specified and linguistic, but that there are no linguistic features specifying location. Liddell contrasts this situation with that seen in spoken languages:

> A word in a spoken language can be described in terms of a discrete set of features which combine to describe the sound structure of that word.... The same cannot be said for signs which use space. I have suggested that ASL and, by implication, other sign languages have classes of signs which can be directed toward elements of grounded mental spaces. Since the elements of these spaces can be located in what appears to be an unlimited number of places, sign languages must rely on a pointing gesture as a component of these signs. This makes such signs unlike words in spoken languages.

Liddell goes on to point out that even though such forms are unlike the words in spoken languages, in both signed and spoken languages people must indicate the referent of deictic constructions. Thus, when a person says "This is the one I like" it must be accompanied by some sort of deictic gesture, such as a pointing finger or a head nod in the direction of the referent. Liddell (2000: 254) claims that these gestures are easily distinguished from the linguistic signal in orally produced languages. The same function in signed languages results in the creation of a set of signs, Liddell claims, which "combine sign and gesture" (354). Thus, Liddell suggests that location in all of these forms is a non-linguistic, gestural element, and that these forms are mixtures of linguistic and non-linguistic, gestural elements (1995:26).

In his later work (Liddell 2003), Liddell no longer specifically talks about a tripartite structure of real, surrogate, and token space. Instead, his framework relies on real space and blended spaces of various types. Liddell also makes an important claim about gradience in signed languages. Signed languages are different than spoken languages, he says, because it is impossible to address their structure without dealing with gradient properties – which we have seen he classifies as non-linguistic and gestural – whereas these properties have been largely ignored in spoken languages. Thus, as in his previous work, Liddell sees this as a way of addressing how gestural and gradient properties interact with the grammatical properties in the process of utterance generation. We will explore other perspectives on this relationship between language and gesture in Section 1.3.5.

Paul Dudis has extended the blended spaces analysis of ASL with his recent work on body partitioning in ASL narrative. Dudis (2004) examines the partitionable zones of the body used during the articulation of signed utterances, such as hands, arms, and even parts of the face. He points out that these body subparts can participate in mappings including two input mental spaces that participate in a blend. One of the inputs is real space; the other input is a narrative space containing elements introduced during a narrative. For example, Dudis demonstrates how a conceptual integration network (Fauconnier & Turner 1998) is produced during a signed description of someone getting punched in the face. One way to sign this would be for the signer to map the assailant role onto herself, creating a visible assailant in real space, and to throw the punch forward towards a nonvisible victim. Body partitioning permits another way, however. Here the punch is described from the victim's perspective.[3] In this case the victim maps onto the signer, creating a visible victim; the assailant is invisible until the actual blow takes place. The signer maps the assailant's forearm onto her partitionable forearm, creating a visible assailant's forearm (see Figure 3). When the blow takes place, the blended space consists of the victim's chin (represented by the signer's chin) along with assailant's partitionable forearm and fist (represented by the signer's forearm and fist).

2.4 Gesture and grammaticization

In the past whenever signed languages have come to the attention of language scholars, they were often described as either elaborate systems of gestures, essentially non-linguistic, or as mere substitutes for speech. The driving motivation of signed language research in its early years was to distinguish signed languages from gesture. Even today, most signed language scholars bristle at suggestions that signed languages may have any connection to gesture. This is not surprising given the intellectual history that has surrounded the role of signed languages in deaf education. Consider, for example, the following statement from

---

3. Although he notes the connection, Dudis leaves unexplored the relation between this option and passive constructions in ASL (Janzen et al. 2001).

Figure 3. Blended space, in which the signer's head represents the victim and the right forearm represents the assailant

Giulio Tarra made during the infamous 1880 Milan Conference when hearing educators rejected signed languages in favor of speech for educating deaf children:

> Gesture is not the true language of man which suits the dignity of his nature. Gesture, instead of addressing the mind, addresses the imagination and the senses. Moreover, it is not and never will be the language of society … Thus, for us it is an absolute necessity to prohibit that language and to replace it with living speech, the only instrument of human thought. … Oral speech is the sole power that can rekindle the light God breathed into man when, giving him a soul in a corporeal body, he gave him also a means of understanding, of conceiving, and of expressing himself. … While, on the one hand, mimic signs are not sufficient to express the fullness of thought, on the other they enhance and glorify fantasy and all the faculties of the sense of imagination. … The fantastic language

> of signs exalts the senses and foments the passions, whereas speech elevates the mind much more naturally, with calm and truth and avoids the danger of exaggerating the sentiment expressed and provoking harmful mental impressions. (Lane 1984: 391, 393–394)

Given such a view, and Tarra's is not uncommon even to this day, it is little wonder that serious scholars have devoted their energy to establishing the linguistic status of signed languages. This often has meant rejecting any claim that signed languages are related to gesture.

The tide is turning, however, and increasingly we see linguists willing to explore the connection between gesture and signed languages (Armstrong, Stokoe, & Wilcox 1995; Liddell 2003; Wilcox 2004b). For the most part, this work takes place within a cognitive linguistic framework.

Two broad areas of empirical study informed by cognitive linguistic and functional frameworks are now being conducted on the relation between gesture and signed languages. The first area looks to gesture as a source of lexical signs. Janzen (1998) suggests that topic markers in ASL may have a gestural source in the expression of surprise. Wilcox and Wilcox (1995) identify gestural sources of modal forms in ASL. Their work is extended in Shaffer (2000, 2002) and in Janzen and Shaffer (2002).

This work actually began as the study of grammaticization in signed languages. Grammatizication is the linguistic process by which grammatical material (for example, grammatical morphemes such as future markers or modal auxiliaries) develops historically out of lexical material (Bybee et al. 1994). Grammaticization operates much the same in signed languages such as ASL as in spoken languages. For example, modals in ASL (indicators of obligation, necessity, or markers or the speaker's degree of commitment to the truth of a proposition) develop historically out of lexical material with concrete, embodied meanings (S. Wilcox & P. Wilcox 1995; Shaffer 2000). Thus, the ASL modal auxiliary meaning *can*, even when used for mental ability (as in *He can read*), developed historically from the ASL sign STRONG 'having physical strength'.

Similarly, Wilcox and Wilcox (1995) describe the development of the lexical sign BRIGHT into the modal form OBVIOUS. Shaffer describes a set of grammaticization paths leading from lexical to agent-oriented, and in some cases to epistemic modal forms. These researchers noted that in many cases gestural sources of the lexical forms can be identified: for example, STRONG appears to have a gestural source in a depiction of upper body strength, and BRIGHT in the depiction of emanating rays of light. Wilcox (2004b) describes the same path of development from gestural to lexical to grammatical in Catalan Sign Language (LSC). The LSC lexical sign PRESENTIR, denoting the sense of smell, may also be used in a more grammatical sense to express the speaker's inferences about actions or intentions, as in

(1) She said she wouldn't go to Holland, but I feel she'll change her mind

where the phrase "I feel she'll change her mind" would use PRESENTIR, literally "it smells like she'll change her mind." PRESENTIR has a gestural origin which indexes the nose; it is a two part sign, made first by pointing to the nose with the dominant hand, and then rubbing together the thumb and fingertips of both hands.

Another example comes from the development of the agentive suffix -PERSON (similar to the English *-er*) in ASL. Historically, this suffix originated as a full lexical form meaning *body*. Over time, the orientation and location parameters changed and the movement became greatly reduced. Semantically, the sign changed from meaning strictly *body* to become *one who does something [as specified by the verb]*. It also became a bound form obligatorily attached to a verb. The current sign *teacher* 'teacher' in ASL is thus the free lexical form TEACH and the reduced suffix -PERSON which developed from the sign for body.

The second area of research to focus on gesture examines how certain gestural properties of signs codify, becoming at first paralinguistic or prosodic, and eventually in certain cases becoming fully grammatical. Wilcox (2004b) claims that manner of movement is such a property. His argument is that manner of movement in certain cases moves from expressing gradient meanings much the same as stress (Wilbur 1999). At this stage, these meanings are akin to what Bolinger (1986: 19) called a "vocalized gesture" such as the delayed release in *I was a f-f-f-fool to do that!* or *It was really b-b-b-big!* Distinct from the spoken forms described above, however, in this case of signed language grammaticization, the path leads from gesture, through prosody, directly to grammar, bypassing any lexical stage. For example, one way in which these forms become incorporated into the grammatical system is as markers of modal strength: weak obligation *should* and strong obligation *must* are expressed with the same base form in ASL, the weak form using a soft reduplicated movement and the strong form a sharp, delayed release movement.

Grammaticization in signed languages is an area of active study, often combining functionalist frameworks compatible with cognitive linguistics with empirical methodologies (Janzen & Shaffer 2003; Meir 2003; Sexton 1999; Shaffer 2000; Zeshan 2003).

2.5 Evolution of language

One unique area in which the study of gesture, signed languages, and cognitive linguistics come together is in the renewed scientific interest in the evolution of language. For centuries, the origins of language remained a purely speculative question, so much so that in 1866 the Linguistic Society of Paris imposed a ban on its discussion (Armstrong 1999). With the advent of cognitive science and neuroscience studies of the brain, advances in the empirical study of animal vocal and gestural communication systems, and the cognitive linguistic approach to language which does not require a categorical distinction between language and other cognitive systems, the study of the evolution of language has moved into a new era of empirical research.

Ever since its earliest, speculative days, explanations of the origins of language have been divided along two assumptions. The first assumes that language is to be equated with speech; the place of signed languages in such explanations is resolved either by assuming that they are speech surrogates or that some highly abstract, modality-neutral system of grammar accounts for both spoken and signed languages.

The second explanation assumes that language began as gesture. For example, gestural primacy theory posits that the earliest proto-language was visual-manual, with a later shift to auditory-vocal. Hewes (1996: 582–583) summarizes these two explanations:

> There are only two major models for the emergence of language ... The first major hypothesis holds that language originated in the vocal-auditory channel as a refinement of the call-systems already present for social communication in the primate and other mammals. The other is that the earliest medium for language lay in gesture, with a later shift to the vocal medium. (1996: 582–583)

Armstrong and his colleages (Armstrong, Stokoe, & Wilcox 1994) have proposed an alternative gestural theory which assumes that all language, and indeed much animal communication, is essentially gestural, in that it is based on kinesic actions produced by moving parts of an animal's body, resulting in acoustic and optical signals. These signals, whether intentional or not, and whether intended to be communicative or not, are interpreted by others as meaningful (Wilcox 2004b).

Hewes (1996) argued convincingly for the primacy of gesture. Part of his argument consisted in questioning the suitability of sound as the original basis of language: "Sounds convey much useful information to primates, but, on the whole, less than is carried on the visual channel" (583). Another aspect of his claim rested in the importance of manual motor control and the cerebral lateralization of language: "If the earliest language or proto-language had been gestural rather than vocal, it could well be that its subsequent striking left-lateralization in the brain resulted from language's having been a function tacked on to the kinds of precise manual manipulations involved in the predominantly right-handed making and using of tools" (584).

Although few people realize it, from his earliest writing Stokoe was interested in examining the relation between gesture and signed language (1960). He claimed that, assuming material culture, familial patterns, and attitudes towards life and "the normal" had advanced to a degree that would permit deaf children to survive:

> there would grow up between the child and those around a communicative system derived in part from the visible parts of the paralinguistic but much more from the kinesic or nonverbal communication behavior of that culture. ... To take a hypothetical example, a shoulder shrug, which for most speakers accompanied a certain vocal utterance, might be a movement so slight as to be outside the awareness of most speakers; but to the deaf person, the shrug is unaccompanied by anything perceptible except a predictable set of circumstances and response; in short, it has a definite meaning" (1)

Stokoe (ibid.: 1–2) also clearly saw the relation of this discussion to a gestural theory of the origin of language:

> Communication by a system of gestures is not an exclusively human activity, so that in a broad sense of the term, sign language is as old as the race itself... This hypothetical discussion of the origin and development of the gesture language of a congenitally deaf individual in any society is not to be taken as a prejudgement of the vexed question of language genesis. ... The kinesic, or more broadly, the metalinguistic communicative phenomena out of which the primary communicative patterns of the deaf are built may once have been the prime sign phenomena, with vocal sounds a very minor part of the complex.

Over the course of the next 30 years, Stokoe developed his thoughts on signed language, gesture, and the evolution of language into a framework that he called 'semantic phonology'. Simply stated, semantic phonology captures the fact that the articulators of signed languages, hands moving through three-dimensional space, carry with them inherent conceptual properties (Stokoe 1991: 111–112):

> What I propose is not complicated at all; it is dead simple to begin with. I call it semantic phonology. It invites one to look at a sign ... as simply a marriage of a noun and a verb. ... [O]ne needs only to think of a sign as something that acts together with its action. ... The usual way of conceiving of the structure of language is linear: First there are the sounds (phonology), these are put together to make the words and their classes (morphology), the words in turn, are found to be of various classes, and these are used to form phrase structures (syntax), and finally, the phrase structures, after lexical replacement of their symbols, yield meaning (semantics). A semantic phonology ties the last step to the first, making a seamless circuit of this progression. The metaphor for semantic phonology that jumps to mind is the Möbius strip: the input is the output, with a twist.

Semantic phonology suggests that visible gestures, whether the common everyday gestures hearing people make when they speak or the conventionalized gestures that are the signs of natural signed languages, are primordial examples of self-symbolization. The twist in the Möbius strip is that the phonological pole of gestures and signs consists of something that acts and its action. Hands are objects which move about and interact energetically with other objects. Hands are thus prototypical nouns, and their actions are prototypical verbs. Hands and their movements manifest archetypal grammatical roles. Cognitive iconicity, discussed above, is simply the cognitive linguistic extension of Stokoe's notion of semantic phonology, raising the question of what role cognitive iconicity – conceptual mappings across phonological and semantic spaces – may have played in the evolution of language.

Cognitive linguists are just now beginning to weigh in on the issue of the evolution of language. Fauconnier and Turner (2002), for example, devote a chapter to exploration of the role that mental space blends might have played in the evolution of language. Their proposal for the origin of language points to a deep, underlying cause, "namely, the continuous development of blending capacity until it arrived at the critical point of double-scope blending" (Fauconnier & Turner 2002: 186). According to their model (ibid.: 131):

> A double-scope network has inputs with different (and often clashing) organizational frames as well as an organizing frame for the blend that includes parts of each of those frames and has emergent structure of its own. In such networks, both organizing frames make central contributions to the blend, and their sharp differences offer the possibility of rich clashes.

They go on to identify the real-world scenes which may have fostered selection of the ability to perform double-scope blending (ibid.: 180):

> There are many scenes that are immediately apprehensible to human beings: throwing a stone in a particular direction, breaking open a nut to get the meat, grabbing an object ... Double-scope blending gives us the supremely valuable, perhaps species-defining cognitive instrument of anchoring other meanings in a highly compressed blend that is like these immediately apprehensible basic human scenes, often because such scenes are used to help frame the blend.

It is interesting to consider how gestural theories of language origins, cognitive iconicity, signed languages, and the proposal that double-scope blending played a critical role in the emergence of language might be integrated into a more encompassing cognitive view of language evolution. Questions that can be addressed by further empirical research include:

– What role did hand actions play in the evolution of language? Fauconnier and Turner (2002: 376) note the importance of human-scale constructions for space, force, and motion in the development of the cognitive ability to perform blends. They go on to state that "Human action, with motion and intentionality in physical space and time, is a basic human-scale structure" (ibid.: 378). Does the close link between manual activity and language (Arbib & Rizzolatti 1996; Kimura 1993; Rizzolatti & Arbib 1998), and the necessity of grounding of double-scope blending at a human scale, suggest that intentionally moving the hand was a critical human action that led to linguistic conceptualization?
– Even the casual reader of cognitive linguistics cannot help but be struck by the pervasive use of visual perception in accounting for the cognitive abilities that underlie human language, especially construal operations so important to grammar. One may wonder whether this is merely an expository device, or whether in fact visual perceptual abilities are central to language. If they are, what does this suggest for the role of visible actions in the emergence of language?
– Might semantic phonology, as manifest through the pervasive cognitive iconicity found in signed languages, be an archetypal example of double-scope blending?

Obviously the cognitive linguistic exploration of the evolution of language is just beginning. Even at this early stage, however, it is clear that there are numerous ways in which the cognitive linguistic study of gesture and signed languages can inform an empirical approach to the study of the evolution of language.

3. Empirical methods in signed language research

The use of a cognitive framework has been very fruitful for linguistic research on signed languages, as is attested by many of the studies reported in the last section. One of the great advantages of cognitive linguistic approaches, in our view, is the grounding of theory in data. The broad range of cognitive linguistics methods covered in this book can be applied to signed languages as well, but have not, for the most part, been applied by researchers working within a cognitive linguistics framework. With the exception of several recent and upcoming dissertations (Shaffer 2000; Naughton 2002; Dudis 2004; Maroney 2004), the vast majority of work currently being carried out on signed languages by cognitive linguists continues to use the more traditional approach of supporting arguments with the analysis of illustrative examples. The good news for cognitive linguists interested in working on signed languages is that empirical methods have been used within a variety of other theoretical frameworks, setting the groundwork for the adoption of these methods by cognitive linguists. In this final section, we describe a number of methodological issues, as well as outlining some specific methods that promise to be fruitful to cognitive linguists. Where appropriate, we also outline some specific studies that used the methods described.

3.1 Recording and storage of signed language data

All researchers interested in empirical studies of signed languages must face the complications involved in recording and storing a visual rather than an auditory signal. Two issues in particular can be difficult to manage: confidentiality of research participants, and memory demands for video storage and manipulation.

3.1.1 *Confidentiality*

The past two decades have seen a sudden rise in governments' recognition of the need to protect the rights of research participants (for further information on the use of human subjects, please see Gonzalez-Marquez et al. this volume). One of those rights is the right to confidentiality. An individual's consent to participate in research does not imply that the researcher may divulge their participation. Spoken language corpora are often shared across researchers (e.g., Switchboard, TalkBank, etc.) because an audio recording of the human voice is not sufficient for identifying an individual uniquely. It is simple, by contrast, to uniquely identify an individual on the basis of a video recording. Thus, signed language data cannot be shared directly without prior permission of the research participants. If explicit consent to make data public is not obtained, video recordings must be modified to remove identifying features, for example, by blurring or morphing facial features, or the data must be transcribed before they can be shared. The first of these options is technologically complex and time-consuming; the second option is complicated by the lack of standardization of transcription methods for signed languages. Thus, to date, there are few public-domain sign language corpora (but see www.bu.edu/asllrp).

3.1.2 *Memory demands*

Video data storage requires much more memory than audio data storage. This issue affects nearly every type of empirical method. There is less historical data available because the technology to record video was developed later than the technology to record audio. It is still inconceivable to generate signed language corpora that would be comparable to spoken language corpora in use in cognitive linguistics, containing hundreds of thousands of utterances, because the memory demands are so great. Experiments using language stimuli run more slowly because opening and closing video files takes much longer than audio files. There are also fewer programs designed to support video in research contexts. As with the issue of confidentiality, one obvious solution to memory issues is to transcribe data into a text format. This reduces the memory demands for storage, and also makes it easier to search the data for specific items or patterns. However, this again raises the question of the standardization of transcription methods.

3.2 Transcription

Currently there is not a standard form of notation for transcribing signed language data. We trace this problem in large part to the fact that there is no widely used writing system for signed languages. Researchers most interested in phonetic or phonological questions have had to generate novel symbols to represent the segments of signed utterances, as seen for example in Stokoe notation (Stokoe et al. 1965), SignFont (Newkirk 1987), HamNoSys (Prillwitz & Vollhaber 1989), and Sutton SignWriting (Sutton 1981). These notation systems are for the most part phonetic, including symbols to represent handshapes and movement, and in some cases, location. Unfortunately, the symbols are different for each system. With the exception of SignWriting, each of these notation systems has a set of conventions for ordering the phonetic symbols in a linear order; however, this convention differs from one system to another (Miller 2001). SignWriting places the handshape and movement symbols in a spatial configuration with a symbol for the head to encode locative parameters instead of using a linear sequence of phonetic symbols. This approach is intuitive for new learners, but nevertheless requires mastery of a set of conventions – different from those of the other transcription systems – to become a proficient reader or writer. (See Miller 2001, for discussion of additional differences in phonetic transcription systems.) A final problem with using any of these notation systems is that they require symbols that are not available on standard computer keyboards, with the exception of SignFont, the least wide-spread of these transcription systems.

Researchers who are primarily interested in morphology, syntax and discourse typically avoid phonetic transcription systems and transcribe data with glosses accompanied by additional markers of facial expression, reduplication, agreement, and so forth. A gloss is a translation of the target sign into a language that has a widely used writing system. One wide-spread convention is to write glosses in capital letters, for example, to transcribe a sign expressing the concept "clean" in ASL, a researcher might use the gloss CLEAN. Glossing is problematic for several reasons (cf. Pizzuto & Pietrandrea 2001). For one, meaning is mapped onto the lexicon differently in every language. Thus, one sign expressing the concept *clean* in ASL also expresses the concept *nice*, which is not apparent from the gloss CLEAN. Alternatively, some English words, like "right" have multiple meanings that are expressed by different signs in ASL, making the gloss RIGHT ambiguous. Thus, the lexicon of the language used for glossing influences the semantic connotations of the transcription since readers typically cannot suppress activation of that lexicon while reading the transcription. Finally, although the use of glosses is widespread, the conventions governing the use of additional markers varies widely. Compare the transcriptions in (2) and (3) of the same utterance in ASL.

(2) PRO.3 BUY[I:iterative] t_i YESTERDAY W-H-A-T_i

(3) PRO-3 PURCHASE+ YESTERDAY $_{KB}$**fs-WHAT**$_{KB}$
What was he buying all day yesterday?

There have been several domain-specific attempts to regularize transcription of signed languages. Neidle and associates have generated a program for tagging video called SignStream which is accompanied by a transcription system, and is distributed free of charge.

SignStream was developed primarily for descriptive analyses of signed languages. Slobin and colleagues have developed a transcription system (Berkeley Transcription System, BTS) for signed languages that will interface with the Child Language Database Exchange System (CHILDES; MacWhinney 1978), which is used primarily by investigators interested in language acquisition. Hoiting & Slobin (2002) explain that the smallest unit of transcription in BTS is the morpheme, because this is currently what is theoretically most interesting and pressing about acquisition research. As Miller (2001) points out, every transcription system rests on certain theoretical assumptions. Thus, as transcription becomes more standard, it is important that the systems be flexible enough to grow with the field and to be modified to fit the goals of the research being pursued. We turn now to a discussion of several specific empirical approaches available to signed language researchers.

3.3 Historical

For many years, historical research on signed languages was shunned, perhaps because when signed languages were regarded as merely gestures and not true languages, it was common to see informal descriptions of signs purporting to show where they came from. These descriptions were in most cases incorrect and were felt by many native users to denigrate their language.

The first formal historical research on signed languages appeared in the sociolinguistic literature, as linguists such as Woodward and De Santis (1977; Woodward 1979) explored the historical linkages between ASL and LSF. Today, with more interest in issues such as grammaticization, historical studies of signed languages are once again being conducted. For cognitive linguists, this is important because such studies can shed light on the cognitive processes that play a role in lexicalization, grammaticization, and language change generally. For example, Heine and his colleagues claim that metaphor plays a central role in grammaticization, while Traugott and her colleagues find that metonymy is important in semantic change.

A number of sources are available for studying historical data. When studying ASL, for example, these include:

- Historical films now transferred to videotape, available from Sign Media Inc.
- A database of historical films including transcription under development by Supalla (2001)
- Old dictionaries and glossaries of signs, such as Long (1918) and Higgins (1923)
- Historical documents describing early ASL, which can be found in such sources as the American Annals of the Deaf
- Archival sources, such as the archives at Gallaudet University or the library at the Institut National de Jeunes Sourds in Paris.

Of course, relying on such historical documentation is fraught with danger. Most of the written accounts are by non-specialists, and many are by authors whose knowledge of the language may be questionable. They are by their very nature documents of an unwritten language and so, unless they are accompanied by drawings or photographs, the researcher has to rely on written descriptions to determine the form of the sign, and glosses to determine lexical meaning. When grammatical information is given, there is usually a strong intrusion from the surrounding spoken language: ASL signs, for example, are assigned to the grammatical categories of their English translation equivalents. Typically very little if any usage information is provided, even in so-called dictionaries such as Long (1918).

3.4 Survey and questionnaire

One empirical method used successfully for many years by functional linguists is the survey. Surveys seem to be used less among cognitive linguists, and are only now beginning to be used by signed language linguists. Although not a survey, the descriptive studies questionnaire designed by Comrie and Smith (1977) is a guide for collecting and describing data. It has the advantage that quite a few typological studies of spoken languages have been conducted using it. The questionnaire outlines broad areas of language structure, such as nominal, verbal, complements, negation, modality, and so forth. Researchers do not need to use the entire questionnaire, but can rely on it to structure a specific area of study.

The Dahl survey is more specific, focusing on only on the domain of tense and aspect. It is a true survey, however, and presents a complete range of questions and scenarios for collecting data. Maroney (2004) has used a portion of this survey to investigate the expression of aspect in ASL.

In her study of negatives and interrogatives in 38 signed languages, Zeshan (2004a, 2004b) reported that she used a questionnaire which also often included visual data in the form of pictures or video. Her discussion of the methodological problems inherent in conducting typological research on signed languages, especially in Zeshan (2004b), is particularly useful.

As more cross-linguistic and typological studies on signed languages are conducted it seems likely that some type of survey format, perhaps modified to permit collection of video data, will become increasingly important.

3.5 Corpus

Corpus studies are a valuable empirical method in functional linguistics and are being used more and more by cognitive linguists (see for example, Kemmer & Barlow 1996). The field of signed language linguistics has seen very little use of corpora. This is understandable, given the severe limitations on collection and transcription described above.

Nevertheless, linguists working on signed languages can and should consider small scale corpus studies when appropriate. Increasingly, computer and video technology is making it possible to collect, store, tag, and search video data using database applications such as 4D, FileMaker, and MySQL.

Studies which include a corpus as an important part of their methodology include Bayley, Lucas and colleagues' studies of variation in ASL (Bayley, Lucas, & Rose 2000; Bayley, Lucas, & Rose 2002; Lucas, Bayley, Reed, & Wulf 2001; Lucas, Bayley, Rose, & Wulf 2002; Wulf, Dudis, Bayley, & Lucas 2002) as well as Morford and MacFarlanes's (2003) study of sign frequency in ASL. Although signed language corpus studies have not yet risen to the size and sophistication that spoken language studies such as those based on the Switchboard corpus or the CHILDES database have, this remains an increasingly viable method which cognitive linguists should adopt for signed language research.

3.6 Experimental

Laboratory investigations of signed languages played a large part in moving the scientific community toward the recognition that signed languages are languages. Thus, research on signed languages is deeply rooted in empirical methods, particularly experimental approaches. One of the earliest groups to investigate ASL systematically consisted of Ursula Bellugi, Edward Klima and colleagues at the Salk Institute. The focus of their research has been psycholinguistic in nature since its inception, trying to uncover the relationship between sign language and the brain, including understanding the effects of modality on language structure, on language processing and acquisition, on cognition more generally, and on brain organization.

One example will illustrate how experimental studies of sign language usage contribute to our understanding of the cognitive basis of language. Some of the early studies carried out at the Salk Institute provided psycholinguistic evidence for the phonological structure of ASL described by Stokoe (1960). If the phonological parameters of handshape, location and movement are used in the representation and processing of signs, then it should be possible to influence signers' behavior by manipulating these elements of signs. In one experiment, Bellugi, Klima and Siple (1975) asked signers to recall lists of signs, and found that errors were remarkably similar to the target signs in phonological form, suggesting that signers hold signs in memory by "rehearsing" the form of the sign. Subsequent studies verified this interpretation by showing that signers made more errors when recalling lists of phonologically-similar signs just as speakers make more errors when recalling lists of words with similar phonological structure (Klima & Bellugi 1979). These studies demonstrated the psychological validity of handshape, location and movement as organizing parameters of signs.

In a similar vein, Lane, Boyes-Braem and Bellugi (1976: 265) carried out a perceptual discrimination study of ASL handshapes in order to "determine what sort of feature analysis for ASL might result if we proceed from psychological data to a linguistic model, rather than the reverse." They presented nonce signs to participants in varying signal-to-noise ratios, and asked participants to identify the handshape of the signs (cf. Miller & Nicely 1955, for English consonants). They were able to propose 11 phonological features that are important for distinguishing handshapes by analyzing the confusability of similar handshapes. Subsequent research has indicated that just three of these original 11 features are sufficient to account for the variability in handshape identification and discrimination abilities (Stungis 1981). While these psycholinguistic studies have contributed to sign language linguistics by providing alternative evidence for the phonological structure of signs, they also demonstrate that phenomena previously observed only in spoken languages are also observed in languages in a completely separate modality. Thus, the whole conception of "phonology" as concerning the "sound" structure of language must be modified to ad-mit to a broader range of evidence. We now know that this level of structure in languages is not specific to sound per se, but to the form of words and signs. This is the fundamental type of insight that can be gained through the study of signed languages.

Experimental methods do have their drawbacks. For one, this approach to the study of signed languages has lead investigators to focus their lens on a tiny sub-population within the Deaf and signing communities. Deaf individuals are highly variable in language experience, much more so than hearing individuals. Since minimizing variability is necessary for experimental research, laboratory studies of sign language behavior have tried to "control" this variability by focusing on native signers. Some work has also compared the signing of native and non-native signers, where non-native signers are often defined as having acquired sign after puberty. There are very few studies that give us insight into signing as used by the vast majority of signers – deaf individuals who were exposed to some form of signing (perhaps not a signed language) in pre-school or kindergarten, and thus have signed for the majority of their lives, and who have hearing families who may or may not have learned to sign as well. These individuals are often highly proficient signers, but are rarely included in experimental investigations of sign language usage. One challenge to be faced by sign language researchers in the future is learning to develop more inclusive methods of experimentation in order to provide models of sign language usage and representation that can be generalized to a majority of signers.

3.7 Narrative

Quite a few studies of signed languages are now relying on various methods for eliciting narratives. One popular elicitation tool is the wordless picture book "Frog, Where are you?" by Mercer Mayer (1969). Participants are asked to look at the book, and then retell the story in their native signed or spoken language. This method has several advantages over other narrative tasks. Most importantly, the pictures provide some insight into the conceptual input to the narrative process so that stories can be compared across individuals who use the same language, for example, to investigate developmental questions (e.g., Bamberg 1987) or across individuals who use different languages to investigate typological differences (e.g., Slobin 1991, 1993), or even to compare narratives in different languages produced by the same speaker (e.g., Silva-Corvalán 1992). An additional advantage of this tool is that cognitive demands on the participants are limited by the presence of the pictures during the narration task. Thus, participants can focus more attention on generating their narratives rather than on dedicating resources to trying to recall the story. There are already several investigations of signed languages that have made use of this tool to elicit data (Galvan 1989, 1999; Galvan & Taub forthcoming; McIntire & Reilly 1996; Morford 1995, 2002, 2003; Taub & Galvan 2001; Wilkinson 2006).

A second method relies on filmed or videotaped stories, again without words. One example is the "Pear Story" film, first used by Wallace Chafe and his colleagues (Chafe 1980). A number of linguists are now using the pear story film to collect cross-linguistic data on signed language narrative. One advantage that the pear story has over the frog story method is that the former is not a children's story. It is presented as live action, and the basic story seems to have a more cross-cultural appeal.

3.8 Multiple methods

It is possible, and in many cases preferable, to combine one or more of the preceding methods when conducting signed language research. For example, historical data from multiple sources may be collected in a database and form the basis for a historical corpus. Since signed language corpora are typically quite limited in scope, surveys and questionnaires can be used to supplement these data. We see the combination of such methods as the vanguard of cognitive linguistics research on signed languages.

4. Conclusion

Cognitive linguistics has brought new energy to studies of signed languages. When applied to signed language research, the insights afforded by this theoretical framework have led researchers to ask new types of questions concerning the nature of language, the cognitive abilities that underlie the human language ability, and how these abilities might have evolved. Although at the present time the use of a broad range of methods to gather empirical data pertaining to these issues is not widespread, it is a new trend in signed language linguistics that must be promoted. The application of empirical methods to cognitive linguistic research on signed languages clearly holds great promise.

References

Arbib, M. A. & Rizzolatti, G. (1996). Neural expectations: A possible evolutionary path from manual skills to language. *Communication & Cognition*, *29*(3–4), 393–424.

Armstrong, D. F. (1999). *Original signs: Gesture, sign and the sources of language*. Washington, DC: Gallaudet University Press.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. (1994). Signs and the origins of syntax. *Current Anthropology, 35,* 349–372.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. (1995). *Gesture and the nature of language.* Cambridge: Cambridge University Press.

Bamberg, M. (1987). *The acquisition of narratives: Learning to use language*. Berlin: Mouton de Gruyter.

Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linkstok Press.

Bayley, R., Lucas, C., & Rose, M. (2000). Variation in American Sign Languages: The Case of DEAF. *Journal of Sociolinguistics, 4*(1), 81–107.

Bayley, R., Lucas, C., & Rose, M. (2002). Phonological Variation in American Sign Language: The Case of 1 Handshape. *Language Variation and Change, 14*(1), 19–53.

Bellugi, U., Klima, E. & Siple, P. (1975). Remembering in signs. *Cognition, 3*, 93–125.

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford, CA: Stanford University Press.

Bouvet, D. (1996). Le corps et la métaphore dans les langues gestuelles: à la recherche du mode de production des signes. [The body and metaphor in gestural languages: In search of the production mode of signs.] In C. Dubuisson & D. Bouchard (Eds), *Spécificités de la recherche linguistique sur les langues signées* [*Specific topics in linguistic research on signed languages*] (pp. 17–34). Montreal: ACFAS.

Brennan, M. (1990). *Word formation in British Sign Language*. Unpublished dissertation, University of Stockholm: Stockholm, Sweden.

Brennan, M. (2005). Conjoining word and image in British Sign Language (BSL): An exploration of metaphorical signs in BSL. *Sign Language Studies, 5*(3), 360–382.

Bybee, J., Perkins, R., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago: University of Chicago Press.

Chafe, W. L. (Ed.). (1980). *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex Publishing Corp.

Comrie, B. & Smith, N. (1977). Lingua Descriptive Studies: Questionnaire. *Lingua 42*, 1–72.

Croft, W. & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Dudis, P. (2004). *Depiction of Events in ASL: Conceptual Integration of Temporal Components*. Unpublished dissertation, University of California Berkeley.

Fauconnier, G. & Turner, M. (1998). Conceptual integration networks. *Cognitive Science, 22*(2), 133–187.

Fauconnier, G. & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York: Basic Books.

Friedman, L. (Ed.). (1977). *On the other hand: New perspectives on American Sign Language*. New York, NY: Academic Press.

Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language, 51*, 676–710.

Galvan, D. (1989). A sensitive period for the acquisition of complex morphology: Evidence from the acquisition of American Sign Language. *Papers and Reports on Child Language Development, 28*, 107–114.

Galvan, D. (1999). Differences in the use of American Sign Language morphology by deaf children: Implications for parents and teachers. *American Annals of the Deaf, 144*(4), 320–324.

Galvan, D. & Taub, S. (forthcoming). The encoding of motion information in American Sign Language. In R. Berman, D. Slobin, S. Stromqvist, & L. Verhoeven (Eds.), *The frog story revisited*.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought.* Cambridge, MA: MIT Press.

Grimes, B. (1996). *Ethnologue*. Dallas, Texas: Summer Institute of Linguistics.

Grote, K. & Linz, E. (2003). The influence of sign language iconicity on semantic conceptualization. In W. G. Müller & O. Fischer (Eds.), *From sign to signing: Iconicity in language and literature 3* (pp. 23–40). Amsterdam: John Benjamins.

Hewes, G. W. (1996). A history of the study of language origins and the gestural primacy hypothesis. In A. Lock & C. R. Peters (Eds.), *Handbook of human symbolic evolution* (pp. 571–95). Oxford: Oxford University Press.

Higgins, D. D. (1923). *How to talk to the deaf.* St. Louis, MO: Catholic church at 1118 N. Grand Blvd.

Hoiting, N. & Slobin, D. I. (2002). Transcription as a tool for understanding: The Berkeley Transcription System for sign language research (BTS). In G. Morgan & B. Woll (Eds.), *Directions in Sign Language Acquisition* (pp. 55–75). Amsterdam: John Benjamins.

Iversen, W., Morford, J., Nuerk, H.-C., & Willmes, K. (2004). Differences between 'six', 'sechs' and '6': A cross-linguistic study of number processing with DGS and ASL. Poster presented at the Eighth Conference on Theoretical Issues in Signed Language Research, Barcelona, Spain.

Iversen, W., Nuerk, H.-C., & Willmes, K. (2004). Do signers think differently? The processing of number parity in deaf participants. *Cortex, 40*(1), 176–178.

Janzen, T. (1998) *Topicality in ASL: Information ordering, constituent structure, and the function of topic marking*. Unpublished dissertation, University of New Mexico: Albuquerque, NM.

Janzen, T. & Shaffer, B. (2002). Gesture as the substrate in the process of ASL grammaticization. In R. Meier, D. Quinto, & K. Cormier (Eds.), *Modality and structure in signed and spoken languages* (pp. 199–223). Cambridge: Cambridge University Press.

Janzen, T. & Shaffer, B. (2003). Implicit versus explicit coding across two languages: Mismatches of cognitive domains during interpretation. Paper presented at the 8th International Cognitive Linguistics Conference. La Rioja, Spain, July 10–23.

Janzen, T., O'Dea, B., & Shaffer, B. (2001). The Construal of Events: Passives in American Sign Language. *Sign Language Studies*, *1*(3), 281–310.

Johnston, T. (2001). Nouns and verbs in Australian sign language: An open and shut case? *Journal of Deaf Studies & Deaf Education, 6*(4), 235–257.

Kemmer, S. & Barlow, M. (1996). Emphatic -Self in Discourse. In A. E. Goldberg (Ed.), *Conceptual Structure, Discourse and Language* (pp. 231–248). Stanford, CA: Center Study Language & Information.

Kimura, D. (1993). *Neuromotor mechanisms in human communication*. Oxford: Oxford University Press.

Klima, E. & Bellugi, U. (1979). *The signs of language*. Cambridge, MA: Harvard University Press.

Kövecses, Z. (2000). *Metaphor and emotion: Language, Culture, and Body in Human Feeling.* Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lane, H. (1984). *When the mind hears: A history of the deaf.* New York, NY: Random House.

Lane, H., Boyes-Braem, P., & Bellugi, U. (1976). Preliminaries to a distinctive feature analysis of American Sign Language, *Cognitive Psychology, 8,* 263–289.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Volume I, Theoretical foundations*. Stanford: Stanford University Press.

Liddell, S. K. (1995). Real, surrogate, and token space: Grammatical consequences in ASL. In K. Emmorey & J. Reilly (Eds.), *Language, gesture, and space* (pp. 19–41). Hillsdale, NJ: Lawrence Erlbaum.

Liddell, S. K. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.), *Language and gesture* (pp. 331–57). Cambridge: Cambridge University Press.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. New York: Cambridge University Press.

Long, J. S. (1918). *The sign language: A manual of signs*. Washington, DC: Gallaudet College Press.

Lucas, C., Bayley, R., Reed, R., & Wulf, A.(2001). Lexical Variation in African American and White Signing. *American Speech*, *76*(4), 339–360.

Lucas, C., Bayley, R., Rose, M., & Wulf, A. (2002). Location Variation in American Sign Language. *Sign Language Studies*, *2*(4), 407–440.

MacWhinney, B. (1978). *The acquisition of morphophonology. Monographs of the Society for Research in Child Development, 43, 1,* 1–123.

Maroney, E. (2004). *Aspect in American Sign Language.* Unpublished University of New Mexico: Albuquerque, New Mexico.

Mayer, M. (1969). *Frog, where are you?* New York: Dial Press.

McIntire, M. L. & Reilly, J. S. (1996). Looking for frogs in the narrative stream: Global and local relations in maternal narratives. *Journal of Narrative & Life History, 6*(1), 65–86.

Meir, I. (2003). Grammaticalization and modality: The emergence of a case-marked pronoun in Israeli Sign Language. *Journal of Linguistics, 39*, 109–140.

Meier, R. (1980). Icons and Morphemes: Models of the Acquisition of Verb Agreement in ASL. *Papers and Reports on Child Language Development*, *20*(Nov), 92–99.

Miller, C. (2001). Some Reflections on the Need for a Common Sign Notation. *Sign Language & Linguistics, 4*(1–2), 11–28.

Miller, G. A. & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America, 27,* 339–352.

Morford, J. P. (1995). How to hunt an iguana: The gestured narratives of non-signing deaf children. In H. Bos & T. Schermer (Eds.), *Sign Language Research 1994: Proceedings of the Fourth European Congress on Sign Language Research in Munich* (pp. 99–115). Hamburg: Signum Press.

Morford, J. P. (2002). The expression of motion events in homesign. *Journal of Sign Language & Linguistics 5*(1), 55–71.

Morford, J. P. (2003). Grammatical development in adolescent first language learners. *Linguistics, 41*(4), 681–721.

Morford, J. P. & McFarlane, J. (2003). Frequency characteristics of American Sign Language. *Sign Language Studies, 3,* 213–225.

Naughton, K. (2002). Linguistic Description and Analysis of Verbs of Visual Perception in American Sign Language (ASL). Unpublished doctoral dissertation, University of New Mexico, Albuquerque, NM.

Newkirk, D. (1987). *SignFont Handbook*. San Diego, CA: Salk Institute and Emerson & Stern Associates.

Pizzuto, E. & Pietrandrea, P. (2001). The notation of signed texts: Open questions and indications for further research. *Sign Language & Linguistics, 4*(1–2), 29–45.

Prillwitz, S. & Vollhaber, T. (1989). *HamNoSys. Hamburg Notation System for Sign Languages. An Introductory Guide*. Columbia, SC: Hornbeam Press.

Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences*, *21*(5), 188–94.

Russo, T. (2004). Iconicity and Productivity in Sign Language Discourse: An Analysis of Three LIS Discourse Registers. *Sign Language Studies, 4*(2), 164–197.

Russo, T. (2005). A crosslinguistic, crosscultural analysis of metaphors in two LIS (Italian Signed Language) registers. *Sign Language Studies, 5*(3), 333–359.

Russo, T., Giurana, R., & Pizzuto, E. (2001). Italian Sign Language (LIS) poetry: Iconic properties and structural regularities. *Sign Language Studies*, *2*(1), 85–112.

Sexton, A. L. (1999). Grammaticalization in American Sign Language. *Language Sciences, 21*, 105–141.

Shaffer, B. (2000). *A syntactic, pragmatic analysis of the expression of necessity and possibility in American Sign Language*. Unpublished University of New Mexico: Albuquerque, New Mexico.

Shaffer, B. (2002). CAN'T: The negation of modal notions in ASL. *Sign Language Studies, 3*(1), 34–53.

Silva-Corvalán, C. (1992). Algunos aspectos de la gramática de los niños bilingües en Los Angeles [Some aspects of the grammar of bilingual children of Los Angeles]. In H. Urrutia Cárdenas & C. Silva-Corvalán (Eds.), *Bilingüismo y adquisición del español* (pp. 227–239). Bilbao, Spain: Instituto Horizonte.

Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics, 1*, 7–26.

Slobin, D. I. (1993). Passives and alternatives in children's narratives in English, Spanish, German and Turkish. In B. Fox & P. J. Hopper (Eds.), *Voice: Form and Function*. Amsterdam: John Benjamins.

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communication systems of the American Deaf, Studies in Linguistics Occasional Paper 8, University of Buffalo.

Stokoe, W. C., Casterline, D., & Croneberg, C. (1965). *A dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet College Press.

Stokoe, W. C. (1991). Semantic phonology. *Sign Language Studies, 71*, 107–14.

Stungis, J. (1981). Identification and discrimination of handshape in American Sign Language. *Perception & Psychophysics, 29*(3), 261–176.

Supalla, T. (2001). Making Historical Sign Language Materials Accessible: A Prototype Database of Early ASL. *Sign Language & Linguistics, 4*(1–2), 285–297.

Sutton, V. (1981). *Sign Writing for everyday use*. Boston, MA: The Sutton Movement Writing Press.

Taub, S. (2001). *Language in the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Taub, S. & Galvan, D. (2001). Patterns of conceptual encoding in ASL motion descriptions. *Sign Language Studies, 1*(2), 175–200.

Valli, C. & Lucas, C. (1995). *Linguistics of American Sign Language*. Washington, DC: Gallaudet University Press.

Wilbur, R. B. (1987). *American Sign Language: Linguistic and applied dimensions*. Boston, MA: College-Hill Press.

Wilbur, R. B. (1999). Stress in ASL: Empirical Evidence and Linguistic Issues. *Language and Speech*, *42*(2–3), 229–50.

Wilcox, P. P. (2000). *Metaphor in American Sign Language*. Washington, DC: Gallaudet University Press.

Wilcox, P. P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics 15*(2), 197–222.

Wilcox, P. & Wilcox, S. (1995). The gestural expression of modality in American Sign Language. In J. Bybee & S. Fleischman (Eds.), *Modality in grammar and discourse*. Amsterdam: John Benjamins Publishing Company.

Wilcox, S. (1992). *The phonetics of fingerspelling*. Amsterdam: John Benjamins.

Wilcox, S. (2002). The iconic mapping of space and time in signed languages. In L. Albertazzi (Ed.), *Unfolding perceptual continua* (pp. 255–81). Amsterdam: John Benjamins.

Wilcox, S., Wilcox, P., & Jarque, M. J. (2004). Mappings in conceptual space: Metonymy, metaphor, and iconicity in two signed languages. *Jezikoslovlje*, *4*(1), 139–56.

Wilcox, S. (2004a). Conceptual spaces and embodied actions: Cognitive iconicity and signed languages. *Cognitive Linguistics*, *15*(2), 119–147.

Wilcox, S. (2004b). Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture, 4*(1), 43–75.

Wilkinson, E. (2006, January). Exploiting location and movement to construct metaphors: A cross-linguistic analysis on American and Italian Sign Languages. Poster presented at the workshop on Cross-linguistic Sign Language Research, Max Planck Institute, Nijmegen.

Woodward, J. (1979). Quelques aspects sociolinguistiques des langues des signes americaine et française [Sociolinguistic aspects of French and American Sign Language]. *Langages, 13*, 78–91.

Woodward, J. & De Santis, S. (1977). Negative incorporation in French and American Sign Language. *Language in Society, 6*(3), 379–388.

Wulf, A., Dudis, P., Bayley, R., & Lucas, C. (2002). Variable Subject Presence in ASL Narratives. *Sign Language Studies, 3*(1), 54–76.

Zeshan, U. (2003). 'Classificatory' constructions in Indo-Pakistani Sign Language: Grammaticalization and lexicalization processes. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 113–141). Mahwah, NJ: Lawrence Erlbaum.

Zeshan, U. (2004a). Negative constructions in sign languages. *Linguistic Typology, 8*(1), 1–58.

Zeshan, U. (2004b). Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language*, *80*(1), 7–39.

Looking at space to study mental spaces

Co-speech gesture as a crucial data source in cognitive linguistics

Eve Sweetser

1. Introduction

The vast majority of linguists, psychologists and cognitive scientists do not look at gestural data. Most such departments offer no training in the transcription or analysis of gesture – and this includes departments that pride themselves on being oriented towards the rich data of embodied cognition and language. Despite this lack of institutionalized curricular niches, research on gesture has made major strides in the last couple of decades. And as a result, widely dispersed individual researchers on language and cognition are finding that the study of normal bimodal language production *as* bimodal, rather than as written transcripts of the auditory modality, gives us far richer data for the study of cognition. The formation of the International Society for Gesture Studies in 2002 and the publication of the journal *Gesture* (starting in 2001) are both results of interaction among this diverse community, and ongoing reinforcements of such interaction.

Real spoken language production in context is always a multimodal process, performed by a gesturing body embedded in a physical setting. Many cognitive linguists now take seriously the embodiment of human thought[1] and language – the idea that our conceptual and linguistic structures could not be as they are if they were not based in human bodily experience of the world. Nonetheless, as a cognitive linguist, I have undergone successive processes of recognizing the degree to which linguistic methodologies fall short of investigating language as embodied. If we linguists are not still focused on written language, we are still focused on the output of the vocal track in spoken language communication, neglecting the visual-gestural track that is co-performed with it.

And yet the visual-gestural medium offers information almost impossible to squeeze into the oral-auditory medium. Vocal production has highly sequential structure (one

1. See the entire cognitive linguistic corpus of work, but notably Lakoff and Johnson 1999, Lakoff and Núñez 2000, Fauconnier and Turner 2002, and the works leading up to them (Fauconnier 1998; Fauconnier & Turner 1996; Turner & Fauconnier 1995). More recent work on mirror neurons supports this embodied viewpoint as well; cf. for example Gallese and Lakoff 2005.

sound at once) and its direct iconic capacity is limited (of course) to sounds and temporal or sequential structure: it can represent a cat's vocalization with the word *meow*, or ongoing length of an activity with repetition such *as she talked on and on and on*. But it cannot directly iconically represent spatial relations, motion, paths, shape, and size,[2] all of which are easily represented in manual and bodily gesture (Taub 2001). Gesture is also more temporally flexible than speech, since the hands (for example) are relatively independent articulators which can perform separate routines simultaneously; add motion of trunk, head, feet, mouth, and eyebrows, and you begin to get an idea of this flexibility. So we would expect that examining gesture might reveal much about embodied conceptual structure which is less obvious in spoken language.

And indeed, looking at spatial cognition and language, McNeill and co-workers have shown that gestural depictions of motion in space, as well as linguistic structures, are systematically different for typologically different languages; researchers on language and space are now making use of gestural data to reach new conclusions about crosscultural aspects of cognitive structure. Cienki (1998) has shown that systematic metaphoric gestural structures may occur separately from linguistic metaphoric usage, as well as accompanying such usage: in such a case, a cognitive semanticist or mental space theorist might want to re-assess the "meaning" of the accompanying language, in the light of the co-produced gestures. Sizemore and Sweetser (in press) show that gesture beyond the speaker's personal gesture space, into the interlocutor's space, happens specifically when the speaker is regulating the conversational exchange; discourse analysts might therefore wish to examine gestural structure alongside linguistic discourse markers. Gesture is as varied in its functions as language, perhaps even more so. But all of these directions, and more, will be of use to cognitive linguists and cognitive scientists.

This chapter has several goals. First, I shall briefly lay out some of the well-established findings which underlie modern gesture research: for example, the evidence that gesture and speech are co-produced as a single neural package, and that a specific pattern of neural packaging is shared by a given language community. This will bring up some basic work in the field – such as that of McNeill (1992, 2000) and Kendon (1990, 2000, 2004). I shall evaluate ways in which such analyses and such data are useful to cognitive linguists.

In the later sections of the chapter, I shall continue to examine the ways in which these tools for gestural analysis can help cognitive linguists to further understand the language and cognition which are going on as the observed gestures are performed – and vice versa. I shall argue in particular that the tools of Mental Spaces theory are extremely productive in this endeavor: gestures should be understood as grounded blends of Real Space (see Liddell 1990, 1995, 1998, 2000, 2003) with other mental spaces. Such an analysis allows us

2. Indirect iconicity is possible as well; for example, sound symbolism systems very frequently involve high front vowels (performed with a felt routine of *small* mouth aperture) representing diminutivity or small size, in opposition to low vowels (with larger aperture) which represent large size (Sapir 1929; Ultan 1978; and see also the papers in Ohala et al. 1994). The vowel sound may be seen as metonymically connected with the size of the aperture involved, which in turn is iconic for the size of the object referred to in the linguistic form.

to see new generalizations at multiple levels of gesture analysis – and also gives us a single framework within which to model multi-modal communication.

Specific examples (presented as photographic frame-freezes from the relevant videos) will be analyzed, showing how gesture operates at abstract levels, expressing metaphoric meanings, discourse-structural meanings, and other abstract conceptual content (ongoing work by the Berkeley Gesture Project will be cited here), as well as iconically representing actual physical spaces. The *systematic* nature of these gestural structures will be laid out, and the ways in which they bear on claims about embodied cognition and mental spaces theory. The goal will be to help readers with differing research goals see the ways in which they can make use of gestural data – and to give them realistic expectations as to the kinds of work involved and the kinds of results which may emerge.

2. Gesture as a back door to cognition

Psycholinguists have become adept at getting below the surface of our conscious language production. Aspects of processing inaccessible to speakers are revealed in minute but readily measurable differences in linguistic processing time, for example, as well as by various scanning techniques revealing neural activity during linguistic processing. But speech-accompanying gesture offers another way of getting at the less conscious aspects of the cognitive processes involved in language. By gesture, I here primarily mean what Kendon calls *gesticulation* – movement of hands and body in flexible visual-gestural patterns which accompany the auditory track of spoken language.

Speakers of all languages gesture as they produce linguistic discourse, constantly. An interesting fact for cognitive science is that co-linguistic gesture is neurally co-processed with language, in tight language-specific neural routines. Stutterers "stutter" in gesture and resume gestural fluency when they resume speech fluency (Mayberry, Jacques, & DeDe 1998). Chinese gestural *strokes* (the main motion phase of a gesture) are systematically co-timed with the object of a sentence (McNeill & Duncan 2000), while English gestural strokes are co-timed with the verb (McNeill 1992). Gestural systems, and the coordination of the gesture-speech package, are acquired over the relatively long time span of language acquisition. Spanish has a different relationship between gestural and linguistic expression of manner and path of motion than English does, and Spanish-speaking children do not seem to have a fully adult system for gesturing about manner of motion until about age 12 (McNeill & Duncan 2000). The gesture-language complex that children learn is learned *as* a complex; coordination between the two tracks is language-specific, and the gestural track is complex and requires lengthy learning alongside the learning of the language system.

Another interesting fact for cognitive scientists and linguists is that the gesture track of this co-production routine is less monitored than the linguistic track. Of course, we cannot consciously choose every linguistic form we produce; we would never express things effectively if much of linguistic production were not automated and processed below the consciously accessible level. (This becomes painfully obvious to an adult learning a second language, where she lacks prefabricated unconscious production routines.) But we are well aware *that* we are producing language (except in cases such as sleep-talking) and we are often aware at intervals of particular form choices ("oh no, I said *men* again, and I should have said *people*"). Speakers notice and may correct the errors of first- and second-language learners: observers do say things like "don't use that word"; "that's *ran*, not *runned*"; "that should be *Do you go*? and not *Go you?*" Admittedly such correction is not common in caregivers of children acquiring language, and second-language learners outside the classroom may also not get much overt correction – and when it does occur it is often ineffectual. But it is easy for an observer to notice such linguistic errors – all speakers are aware of whether a past tense form is correct or not in their dialect. This is much less the case with gesture. Parents in some cultures encourage children to gesture less, or smaller – but they do not notice errors in the form or order of gestures, or suggest alternate better forms. For instance, no researcher has reported a parent saying to a child, "no no, that gesture should have been timed with the subject of the sentence rather than with the verb." And this is unsurprising, given that the parents are themselves unaware of the generalization about gesture timing.

So, as McNeill (1992) and others (see particularly Alibali 1994; Alibali & Goldin-Meadow 1993a, b; Goldin-Meadow 1999, 2003; McNeill 2000) have cogently argued, gesture has immense potential as a source of information about cognitive processes during linguistic production and comprehension. Goldin-Meadow (2003) gives strong evidence that when gesture and speech conflict, two cognitive models are represented. This can be used as a metric of when a child is in the process of catching onto a new concept. For example, Goldin-Meadow observed children responding to Piagetian conservation tasks; at a certain point, when a child has watched the water poured from a tall thin glass to a short fat glass, the child may still *say* that there is less water now – but may gesture a width-wise gesture which suggests awareness of the differences in circumference of the containers. Compared to other children who give the same verbal answer, such a child is more likely to be ready to go beyond a simple metric of height for quantity, and learn the more complex conservation relationship.

Cognitive linguists will naturally want to know whether everyday speakers and gesturers, as well as scientists, can access the information provided by gesture alongside that provided by speech. The answer is pretty clearly *yes*. It does not seem that the cognitive utility of gesture is limited to one side of the communicative exchange, although gesture is differently useful to the performer and the viewer. Speakers gesture even when not in the physical presence of a viewing interlocutor (Bavelas et al. 1992; de Ruiter 1995). Telephone speakers seem to make fewer interactive, discourse-regulating gestures than face-to-face speakers do, although they still make highly routinized interactional gestures such as head-nods. However, they still make iconic gestures about the content of their speech (including spatial gestures while giving directions by phone). All this would seem to indicate that gesture is not there only for the hearer. Krauss et al. (2000) and others have documented the impact of gesture on lexical access; speakers who cannot make manual gestures (e.g., are requested to keep their hands grasping the top of a chair while speaking) have more difficulty with lexical access, particularly when the content of their speech is spatial. This is what a neuroscientist would expect. When two neural routines are closely correlated in performance, activating one of them will help activation of the other (as with humming a tune to help you remember the words of the song). McNeill and Duncan (2000) have posited a more complex model of *growth points*, wherein gesture is a *material carrier* for the dynamic development of thinking for speaking.

Gesture also has communicative impact on hearers/viewers. McNeill (1992) records an experiment where subjects viewed a videotaped narrative that included a gesture-speech mismatch (the videotaped speaker gestured stair-climbing with his fingers while saying *he climbed up the drainpipe*). A significant number of viewers, in retelling the story, said that the character climbed up stairs rather than up a drainpipe. Kendon (1995) presents examples where information is clearly available from the gestural track, complementing the spoken information. Goldin-Meadow (2003) presents an impressive body of evidence that hearers make use of speakers' gestural information. Özyurek (2000) presents interesting evidence of "addressee design" in gesture, and similar patterns are suggested by Sweetser and Sizemore (in press). Perhaps the most impressive evidence of all for the communicative efficacy of gesture, however, are examples like McNeill's (1992) Snow White experiment, wherein subjects were required to tell the story of Snow White without language. The storytellers' communicative partners "took up" the gestural content with remarkable speed, and the speedy reduction and conventionalization of patterns representing ongoing characters give clear evidence of what happens to gesture when it bears the full communicative load unassisted by another track. Goldin-Meadow (1993, 2003) similarly chronicles the use of gesture by deaf children not exposed to a signed language.

Signed languages emerge when gesture is the central vehicle of communication among a community, over a long enough period of time.[3] We have recently been given glimpses of this emergence process in the work of Kegl and colleagues (Kegl & Iwata 1989; Kegl, Senghas, & Coppola 1999) on the development of Nicaraguan Sign Language within a community which began around a Deaf school. Signed languages are thus the endpoint of a communicative development which starts when the visual-gestural modality is regularly given the full communicative load, as spoken languages are the endpoints of a developmental sequence wherein vocalizations ended up bearing the primary semantic load but remained paired with visual-gestural expression. There is, however, good evidence that signed languages also have gesture, performed in the same visual medium as the language performance (Liddell 2003); the same kinds of evidence also suggest that we should be thinking of some components of spoken language prosody as gestural (in *he gave a looooong talk*, the iconic vowel lengthening is gestural). The similarities and differences between spoken and signed language gestural systems are only beginning to be investigated.

I shall also not primarily discuss fully conventional gestures here – what many researchers have called *emblems* and Kendon calls *quotable gestures*. The affirmative head-nod, and the thumb-index circle (ASL "F" handshape) which English speakers use to mean "just right, perfect" are examples; so are various obscene gestures. Basic differences between these gestures and flexible co-speech gesture include (1) that the head nod or the "just right" gesture does not need language at all, but can substitute for words if necessary,

3. See Armstrong, Stokoe and Wilcox (1995) for arguments that such signed languages may have been crucial to the evolution of modern spoken languages.

and (2) that the head nod and the "just right" hand gesture are fully conventional, always formed in the same way and subject to judgments of mis-performance if done wrong. Quotable gestures are complex and well worthy of discussion in their own right; readers are referred to Kendon (1990, 2004); Calbris (1990).

I also do not here deeply discuss all the ways in which bodily location, and in particular the relative bodily locations of interlocutors, are relevant and meaningful to the interaction, or the ways in which gesture indexes all kinds of meanings in discourse. Some of these (particularly Haviland 1993, 2000) will come up in my final discussion of viewpoint. Clark (1996) and Goodwin (2000) are excellent examples of work on these subjects; Engle (2000) and Smith (2003) are fascinating studies of the way different modalities of gesture are used, and of the interaction of gesture with surrounding objects.

And I do not here follow the ways in which gesture has been used as a window on speech processing; this is studied in the work of researchers such as Krauss et al. (2000) and Butterworth and Beattie (1978, 1989). Nor do I adequately cover the ways in which gesture is involved in social interaction; the work of Bavelas (Bavelas et al. 1992 and elsewhere) is of particular interest here. Clearly cognitive linguistics should be interested in all these issues.

This chapter will present some ways that looking at gesture can be useful, in my view, specifically to cognitive linguists. It cannot possibly discuss all the ways. Gesture is an immensely varied phenomenon, and far less studied to date than language. Susan Duncan once said to me that when researchers make conflicting claims about gesture – e.g., that it expresses content, and that it negotiates interaction; or that it facilitates lexical access, and that it is used independently of words – she felt this was usually only evidence that they were looking at different aspects of this pervasive and complex range of human behaviors.

3. Transcription

Every analyst will have different needs, in analyzing and transcribing gesture, just as in analyzing and transcribing language. One linguist may need spectrograms and close phonetic transcription, another may find that standard English orthography is fine for her purposes; another (a discourse analyst perhaps) may not want phonetic transcription but need to measure pauses. Similarly, in studying gesture, there are cases where it is crucial to measure time in milliseconds, as in the research (cited above) leading to our current understanding of the co-timing patterns of gesture and speech, and the evidence for neural co-performance of the two routines. For my own work, it has been sufficient to identify co-timing of a gesture stroke with language at the level of the syllable or word, since my goal has been to examine the relationship between the meanings of the two tracks.

I will therefore not be advocating and presenting in detail any particular gesture transcription system. My own system has been strongly influenced by the McNeill group's work, which uses a grid around the speaker's body for spatial location description, and a modified version of the American Sign Language finger-spelling system for description of handshapes. But most of what you will see in this paper will be relatively abbreviated transcription forms, focusing on the correlations between form and meaning rather than on highly detailed formal description of gesture. Reading McNeill (1992), Kendon (2004) and other works on gesture will provide better resources for someone starting gesture transcription than I can provide here.

It would be yet another paper (and not one for which I am particularly competent), to assess the technical needs of a wide range of possible gesture research projects, including the variation between the needs of researchers with different computer systems. For all gesture analysis and transcription, researchers will need some system of video playback that permits simultaneous slowing down of the visual and sound tracks. Presuming that your data is digital, therefore, Final Cut Pro and other more sophisticated programs will work; QuickTime is not sufficient, although a convenient medium for exchanging video files and for making freezes for illustrations. A transcription program allowing co-indexing of the transcript with the video is also important as the data base grows; needs will vary depending on the project.

I close this section with a final word about the difficulties of data collection and presentation. It should be noted that gesture transcription is much more time-consuming than speech transcription. Like speech, gesture can be transcribed more or less finely; but the standard approximation is that transcribers spend at least ten times as much time on the gesture track as on the corresponding speech track. If you have done speech transcription, you know that this is a daunting figure; and indeed it is correspondingly difficult to build up large corpora of transcribed gesture data. Further, although one can (and linguists often do) present only a transcription of spoken data, it is essential to present photo freezes, if not whole video clips, in presenting analysis of gesture; no transcription system is simultaneously detailed enough and transparent enough to substitute for visual images in a scholarly article. However, since speakers are far more identifiable from visual images than from written transcripts of their words – and yet more from video clips – subjects' privacy is correspondingly more protected, and much analyzed gesture data would be barred from inclusion in any large publicly accessible database.

So linguists thought it was hard to analyze real language data, and they were right; but the complexity and difficulty of the endeavor – and therefore the time commitment – are even greater when you analyze multimodal communication. And of course, you will still be missing a lot, no matter what you choose to focus on – you won't be able to work on every aspect of multimodal communication simultaneously, and researchers with different emphases will see things you do not (just as linguistic researchers notice different things about the same language phenomena). But gesture analysts agree that we can't now go back into the box; communication is multimodal, and we need to work on it as it is.

4. Gesture, iconicity and levels of abstraction

Gesture can, as mentioned above, be literally iconic for a flexible array of physical and spatial entities. Most notably, a body can be iconic for a body: a gesturer who is presenting a represented character's words and simultaneously pounds on the table, for example, can be interpreted as meaning that the quoted character pounded on something. As noticed by Mandel (1977) and Taub (2001) for ASL, iconic mappings between the body and represented content also include:

1. use of some part of the body to represent some object – a finger representing a leg, for example, as a hand gestures "walking" with fingers.
2. representation of an object by representing grasping or other interaction with it – e.g., a hand shaped to grasp a glass, to represent a glass.
3. moving the articulator to trace the shape of an object ("drawing" in the air or on some surface).
4. using motion of an articulator along a path to represent motion of an object along a path.
5. partial representation of some physical routine or activity – for example two hands grasping an invisible steering wheel to represent driving a car (LeBaron & Streeck 2000; this is also close to American Sign Language forms for CAR and DRIVE).

A major problem for iconic interpretation of a representation is always knowing how much to map from one domain to another. For example, in representing a person walking with two fingers pointing downwards, the gesturer and any interpreters have to know that only part of the gesturer's hand is to be mapped –the thumb and two smaller fingers need to be left out of the iconic representation. Dudis (2004a, b) has addressed this problem for ASL iconicity, arguing that signed language involves partitioned *zones* of representation. A parallel solution for gesture seems necessary.

A great deal of work has been done on gesture representing spatial concepts and accompanying spatial linguistic content. There is already solid evidence that cognitive differences correlate with differences in linguistic representation of spatial relations and motion. From Slobin's (1987, 1996, 2000) work on Thinking for Speaking and Levinson's (1997, 2003) work on spatial language and thought, we know that mental images, memory of spatial arrays and other cognitive factors seem to differ depending on whether the speaker's language has an absolute or relative spatial system, and on how the language's grammar combines expression of motion, path and manner. McNeill (1992), Kita (2000, 2003), Özyurek (2000), Nobe (2000), Haviland (1993, 2000) and many others have specifically studied spatial gesture, including the crosslinguistic differences between such gestures.

Equally interesting, however, are gestural representations of abstract concepts. As detailed in Taub (2001) for ASL signs with abstract meanings, and by Parrill and Sweetser (2004) for gesture, such representations are necessarily both iconic and metaphoric. For example, a speaker who shows a clenched fist as she says *rock-solid argument* is iconically representing a solid physical object; but (like the word *rock-solid*) this metaphorically represents an abstract argument which is perhaps likely to withstand counterargument.

In the following examples, transcribed by Fey Parrill and Eve Sweetser and analyzed by the Berkeley Gesture Project, you can see speakers representing abstract structure in terms of location and motion in physical gesture space. The first one is from a public lecture given by Mark Johnson at a bookstore in Los Angeles. Johnson is saying that a traditional understanding of meaning involves semantic concepts with fixed definitions which map neatly onto states of affairs in the world.

...have fixed definitions
and they map onto the world...um...
and that knowledge consists in...framing a set of concepts
that neatly map onto states of affairs in the world
whether those states of affairs have to do with morality or politics or um...or um...
quantum physics or whatever.

As he speaks, he gestures. On *fixed definitions* and *concepts* his hands (palms facing each other, on his upper right) delimit a space for concepts (Figure 1). As he says *map onto the world*, and once again as he says *states of affairs*, his hands move to delimit a space for real-world states, to the lower left of his gesture space (Figure 2). What he is doing is setting up a gestural space wherein the space to his upper right refers to concepts, and the space on his lower left refers to world-states. It is normal for a speaker contrasting two entities to use the two sides of the gesture space in this way; and it is also normal for gesturers to use the vertical dimension as he does (ABSTRACT IS UP, CONCRETE IS DOWN) in contrasting abstract concepts with concrete states of affairs.

Language and gesture show systematic use of the same metaphoric systems (Cienki 1998; Sweetser 1998; Núñez & Sweetser [2006]). Sometimes they are used simultaneously and in parallel – as in the case of the speaker who said *rock-solid evidence* and also made a solid fist gesturally – but gesture can also be used independently of language, as of course language can use metaphor independent of gesture (written language is full of metaphor, after all). Cienki (1998) documents a speaker who consistently gestures downwards for bad grades, up for good grades, and downwards for immoral behavior, upwards for moral behavior. Although the speaker (an American college student) is surely familiar with English linguistic metaphoric uses such as *high moral values*, *low-down dirty behavior*, and *high* or low *grades* (cf. Lakoff & Johnson 1980: 200 he does not use such wording in this particular section of his discourse; only his gestures show clearly that he is conceptualizing morality and grades in terms of vertical scales.

Figure 1.

Figure 2.

Divisions of space in gesture similarly metaphorically represent *mental spaces* or areas of content; spatial loci become associated with subjects (McNeill 1992), as in signed languages (Liddell 2003). Relative positions of loci in gestural space can systematically represent relationships between the content connected with those loci. Just as in linguistic usages such as *far apart* to mean "different" or "socially unconnected" and *really close* to mean "similar" or "socially connected", gesturers bring their hands or fingers into proximity to show similarity and relationship, and make a gesture of moving their hands apart when discussing difference.

The temporal structure of gesture is also exploited iconically to represent temporal structure of represented activities and situations. Dudis (2004a) has given an analysis of the multiple ways in which time is mapped between sign language performance and meaning. Similar structure regularly occurs in gesture. Our next example comes from a public lecture given to the UC-Berkeley Institute of Cognitive Studies colloquium. The speaker's topic is short-term planning memory. At one point he focuses on the "mental book-keeping" involved in spatiomotor tasks: a person dialing a phone number needs to know at what stage in the sequence she is, at any given point, even if she is repeating the same action as an earlier stage.

> ...and a key feature of spatiomotor tasks
> is the book-keeping you need to keep track of where you are
> You dial a phone number, 38537,
> you have to know...
> when you're doing the 3, you have to know where you are,
> is it the first three or the second three?

In this sequence, the speaker consistently gestures outwards and to the right, to show the development of a spatiomotor process; earlier stages are locations closer to his body (and reached earlier, since he's gesturing outwards) and later stages are locations farther from his body, as in the illustrations below. Figure 3 is the starting position (walking around, he

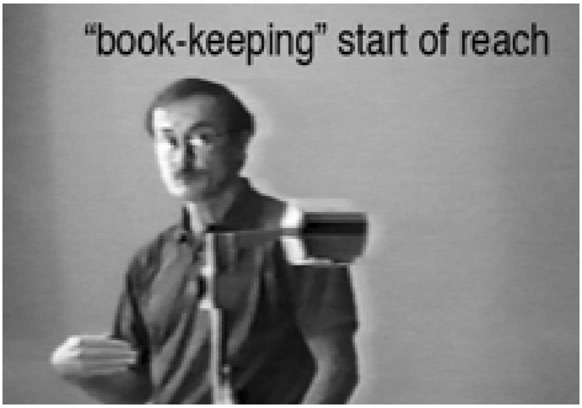


Figure 3.

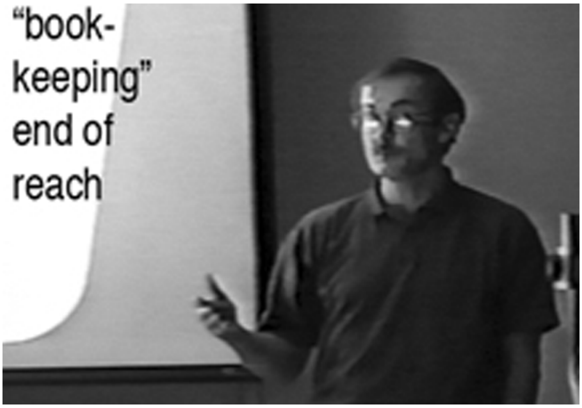


Figure 4.

performs beats in about this position from *book-keeping* to the start of the phone number dialing), and Figure 4 is the final position (on *seven*, the last digit of the phone number). He subsequently retravels this motion as he goes back and forth along the same path from *first three* to *second three*, showing that he has built up a spatial image which is still present for exploitation. Note that although dialing a telephone is a spatiomotor routine, and of course this outwards gesture is a spatiomotor routine too, the gesture does not represent the spatiomotor work of telephone dialing itself; it represents only the temporal sequential structure of the telephone dialing task. It is thus an abstract metaphoric gesture, even though it is about a physical routine.

In general, my observation is that gestural representation of temporally evolving processes, abstract or concrete, necessarily involves motion either *across* (from one side to the other) or *outwards from* the speaker's body (in this case, both). You cannot represent dialing a phone number (or getting a Ph.D., or solving a problem) by motion *towards* the body from a point farther away. There is only one exception, which is that if the process mentioned consists of physical motion towards a deictic center (e.g., if the speaker is saying *he came closer and closer to me*), then it is entirely appropriate to represent this process of approach iconically by moving the hand towards the body.[4] The body seems to constitute a spatial deictic center, which is an origo for the representation of spatial structure – and hence, metaphorically, for the representation of temporal structure as well.

For those interested in embodied cognition, it is important to understand restrictions on the mapping between Real Space (the speaker's perceived gesture space) and mental spaces such as the temporal structure of a process. What is it about our construal of embodied spatial motion – and its correlation with time – which allows our bodies to be the default temporal origo representation? Presumably it is related to the fact that when our muscles are relaxed, our hands fall back to the body, so that extension from the body is in some sense the primary gesture, necessarily preceding retraction.

5. Gesture and levels of content

A single linguistic form is often meaningful not merely in multiple senses, but at multiple levels of the communicative interaction. The labels I use for such levels are *content, epistemic, speech act*, and *metalinguistic*.[5] Gestures have the same multi-level property. Kendon (1995, 2004) documents the use of a single Neapolitan gesture, two hands sweeping in a barrier motion, with discourse meanings as well as content meanings. In one case it accompanies a speaker talking about *stopping* doing something, in another case it seems to indicate that the accompanying linguistic contradiction should cut off (*stop*) the interlocutor's train of reasoning, and in a third case it wordlessly tells someone not to interrupt while the gesturer is engaged in talking to someone else (*stop* the attempt at speech interaction). McNeill (1992) has a discussion of American discourse interactive gestures such as the hand held out, palm upwards, to mean that the speaker feels she has made her point; Sweetser (1998) discusses the American palm-outwards hand used to fend off interruption in conversation.

A crucial point about these gestures is that – being about abstract meanings – they are necessarily metaphoric. They could be used "literally" too; that is, they could in principle iconically represent a physical barrier, or showing a physical object to an interlocutor, or putting out a hand to protect oneself from a blow or a projectile. But from this iconic structure, they are metaphorically mapped onto thought and discourse interaction. Sweetser (1998) points out that the metaphors forming the basis for these abstract interactional

4. Taub (2001) brings up similar instances in ASL grammar, where actual physical motion direction is salient enough to trump other factors in motivating sign structure.

5. The first three of these four concepts were developed in a tradition of scholarship initiated by Halliday and Hasan and are now best known in the work of Traugott (1982, 1989) and Hopper and Traugott (1993) on grammaticalization; my specific terms and definitions were developed in Sweetser (1990). My use of a metalinguistic level stems from the work of Horn (1985, 1989) on negation and Dancygier (1998) on conditionals; it is further developed in Dancygier and Sweetser (2005).

Figure 5.

gestures are also basic linguistic metaphors in English: for example, as detailed in Sweetser (1987, 1992), English regularly treats abstract communication as OBJECT EXCHANGE. Some metaphoric informational "objects", such as hostile questions or unwelcome interruptions, can be seen as projectiles, which in turn can be *rebutted* or *fended off* – iconically represented by the palm-outwards hand in gesture. Objects can be "shown" as well as exchanged; and indeed, a very common linguistic accompaniment to the palm-up gesture is *You see*?, or *See my point*?. The palm-up gesture thus appears to be iconic for showing an object – an *invisible surrogate* object, using Liddell's (1995, 1998, 2000, 2003) and Dudis' (2004b) terminology.

In the following examples, recorded by Marisa Sizemore for the Berkeley Gesture Project, two friends are talking together. S1, the blond woman, is describing a past unsatisfactory roommate to S2, the dark-haired woman. In Figure 5, as she says *clean the underside*, S1 makes a gesture of dishwashing by moving one hand (the "washing" hand) in front of the other (representing the plate being washed). This is a clear example of content gesture, iconic representation of one physical action by another.

S1: She would never clean the undersides of the dis[hes]
S2: [OK],
so she'd wash the dishes
but you had a problem with [how she did it.]
S1: [And I. . .and I. . .]

S2 then interrupts to tease S1 about her fussiness; after all, this roommate did *wash* the dishes. As she interrupts, she reaches far forward into S1's normal *personal gesture space* (Figure 6b) and pats the table on *wash* (Figure 6c); then she leans yet further forward (Figure 6d) as she says *but you had a problem*; she finally leans back again (Figure 6e), satisfied that she has made her point and ready to listen to more of S1's story.

This gesture does not represent anything about the content of the utterance; it shows nothing about cleaning dishes, having problems, or manner of dishwashing. It is, however, a typical interactional regulation gesture. Sweetser and Sizemore (in press) show repeated forwards reaches into an interlocutor's gesture space at moments when the reacher wants

Figure 6a.

Figure 6b.

Figure 6c.

to (re)claim the floor or start up a new topic. Such interactional gestures are typically performed on the line between the speaker and the addressee, between their personal gesture spaces (in what Sweetser and Sizemore call *interpersonal space*). Other kinds of gestures can be performed in *extrapersonal* space, the space outside both personal and interpersonal spaces; for example, at another point in S1 and S2's discourse, S1 reaches far up to her left in an iconic gesture of reaching up for a plate in a cupboard (and finding it to be

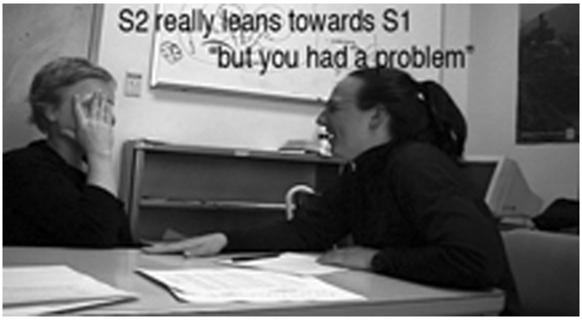


Figure 6d.

Figure 6e.

greasy). It would be unusual, at least, to build the cupboard in the interpersonal space; and it would be impossible to perform floor-claiming in the extrapersonal space.

Gesture thus not only functions at the same multiple levels of communicative meaning as speech, but spatial divisions in the gesturer's Real Space are relevant to functional levels in communication. Physical spaces once more stand for different Mental Spaces. This is of course important for cognitive linguists; and the finding points again to the need to look at all functions of language, not only at content communication.

6. Gesture, viewpoint and deixis

Gesture is performed by a body located and oriented in space, and can exploit the body's location and orientation in multiple ways. Spoken language also does this, of course; in normal face-to-face interaction, references such as *here* can depend on the hearer's knowledge of the speaker's location to index it successfully (a more problematic assumption when *here* is said instead over a cell phone connection). But speech is evidently lacking in indexing precision compared to gesture; *this one* or *that one* may need a supplementary gesture of pointing, but the point does not need a supplementary linguistic *this* or *that*.

Pointing is apparently a human universal, although with many cultural variations in form and use: Kita (2003) gathers recent work on such indexical gestures.

A body necessarily has physical viewpoint (Sweetser 2001, 2002, 2003); human bodies share structure which ensures that they can see forwards but not backwards, can access objects in front of them better than ones in back of them, can move forwards better than backwards, and of course are experiencing a gravitic environment in which we are normally able to stand on our feet rather than our heads (Clark 1973; Fillmore 1997[1971]). As Sweetser (2001, 2002, 2003) has pointed out, other subjective viewpoint phenomena correlate with physical perspective and access to a scene. There are *primary scenes* (Grady 1997, 1998; Johnson 1996, 1999) wherein we co-experience attentional focus, interactional affordances, and all the social and emotional correlates (e.g. satisfaction at successful physical interaction).

So one rather simple mapping is to have a viewpointed body represent another viewpointed body; and this is common in both gesture and signed language structure. It is predictably more conventionalized in ASL; grammarians use the term *role shift* to describe a grammatically distinguished stretch of signed discourse during which the signer ceases eye-contact with the addressee and turns her body slightly to one side, while enacting the linguistic performance and other actions of a "quoted" character. Similar shifts in gesture have been referred to as *character viewpoint* (McNeill 1992).

Global viewpoint has also been observed in gesture, as in signed languages; for example, when a hand represents a moving person, the other hand can represent another person moving to meet the first person. In this case, the gesturer's body trunk, and even the arms, are not representing anything, but are outside the mappings; only the moving hands represent the moving entities.

Rapid switches between these two kinds of viewpoint are apparently quite easy for speakers. The same Institute of Cognitive Studies lecture, cited in Section 4 above, contains such a case. The speaker is talking about reinforcement learning, and is using the metaphor of someone wandering around a maze, "running into" rewards and/or previously encountered locations. Like a rat running through a new maze, the wanderer has no initial direction, but is "wandering around idly". On the first utterance of *wander around idly* (Figures 7a–b) he rotates his right hand in front of his place-holder left hand, showing undirected motion (see Parrill & Sweetser 2004). In this case, only the hand and its motion represent the wanderer in the maze – and indeed the representation is rather abstract, since the general aimless motion is not the kind of motion involved in running around a maze.

Transcript:
you can think of reinforcement learning as dynamic programming done badly…
m'kay… he way reinforcement <Xworking it doesX>
you just wander around idly … m'kay
…and then when you bump into the reward you know what to do from the penultimate state.
You're not gonna have a policy,
you're clueless,
but the penultimate state, you know what to do…

and you know how to <Xdiscount itX> right?
Then you're wandering around idly a-again
you might bump into that penultimate state
the state before that knows how to do...
and you know how to <Xdiscount itX>, right?

However, when he says *you're clueless* and later *you're wandering around idly* again, he waves his arms in mime of a confused, lost person; his whole body represents this person. (The *clueless* image is below as Figure 8, because it is of more reproducible quality than the similar *wandering around idly* image.)

Even when the body is not involved in character viewpoint, it stands as an origo for spatial and temporal representation, as also mentioned above. However, to navigate from an origo one must know what space is being represented. As Haviland (2000) has cogently pointed out, a speaker in (for example) Berkeley may be giving directions to someone about a location in Palo Alto – and in such a case, the pointing and directional gestures

Figure 7a.

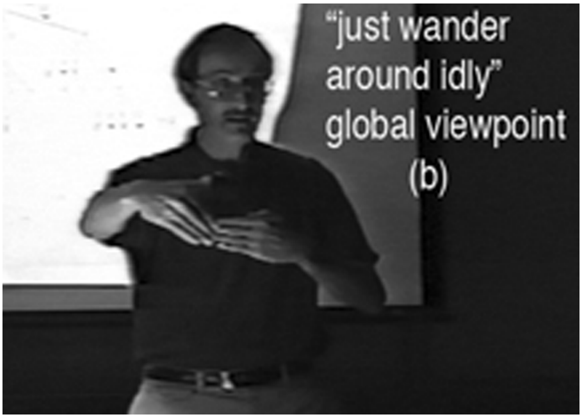


Figure 7b.

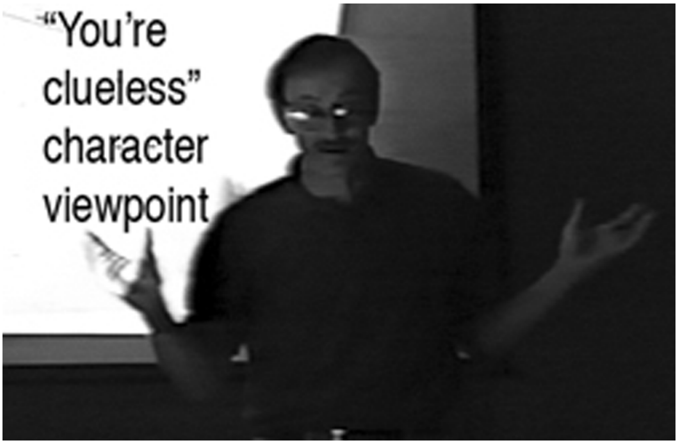


Figure 8.

will not refer to the Real Space of the surrounding city of Berkeley, or to the Real Space north-south and east-west parameters, but rather to imagined surroundings of Palo Alto, assuming an imagined imputed orientation of the speaker with respect to those surroundings. This means that the indexical function of gesture is much richer and more flexible than it appears to be on the surface – although Goodwin (2000) gives excellent examples of how flexible the connections can be between actual objects pointed to, and the intended meaning.

Similarly, deictic temporal gesture necessarily involves use of the body as establishing a spatial origo which is metaphorically interpreted as meaning NOW. Núñez and Sweetser (2006) have shown that in cultures whose languages follow the semi-universal metaphoric models PAST IS BEHIND EGO, FUTURE IS IN FRONT OF EGO, gesture does so too – English speakers clearly gesture forwards about the future and backwards about the past. Aymara, a native American language spoken in the Andean highlands of Chile, Bolivia and Peru, is a genuine exception to this quasi-universal of linguistic metaphor: Aymara linguistically treats PAST as IN FRONT OF EGO, while FUTURE IS BEHIND EGO. As Núñez and Sweetser show, Aymara speakers also gesture forwards when talking about *last year* or *last week*, and backwards when talking about *next year* or *next week*. The absolute universal here seems to be that the front of the body is the NOW origo. (In signed languages also, the NOW location seems to be not the body itself, but the front surface or the space immediately in front of the body). Lateral motion representing time (a speaker can say *it went on for years* and gesture from one side to the other) lacks a specific built-in NOW, and is not deictic in its function.

In Figure 9 below, the speaker from the Cognitive Studies colloquium is finishing his lecture. He says: *[it's time] to..uh for me to stop here and interact with you at this point and and and – take some – take questions and… so I think I'll do that here. Thank you*. As he says *I think I'll do that here* and *thank you*, he twice points downwards to the ground immediately in front of him. Both the word *here* and the point to a location "here" (right in front of Ego) refer metaphorically to NOW, rather than literally to a spatial location.

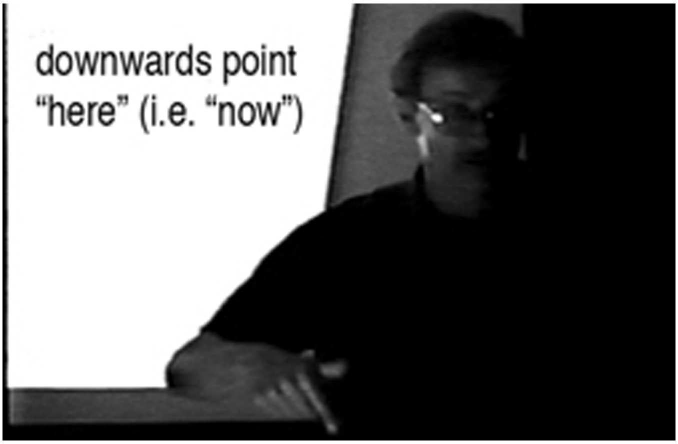


Figure 9.

These points are extremely common in English speakers’ gestures – many a speaker saying *right now* can be seen pointing downwards for emphasis, just as she would if she were saying *right here*.

It would be particularly interesting to have data on the temporal representation gestures used by speakers of languages with *absolute* spatial systems (Levinson 1997, 2003). These speakers do not use their bodies as origos for everyday spatial representation as pervasively as users of relative spatial language systems. An English speaker might say *the ball is in front of the tree* (Hill 1982; Levinson 2003); the hearer would need to know that in English, trees are construed as facing the speaker (Ego-Opposed) and that therefore the ball is between the tree and the speaker. An absolute spatial system would dictate something more like *the ball is west of the tree*, needing no calculation from the speaker’s location. It might therefore generate less egocentric spatialization of time, as well, at the metaphoric level.

7. Conclusions

My work on iconic and metaphoric uses of gesture (Sweetser 1998; Parrill & Sweetser 2004; Núñez & Sweetser 2006), conversation-regulation uses of gesture (Sizemore & Sweetser in press), and general mental space mappings in gesture (Parrill & Sweetser 2004) have convinced me that mental spaces theory is how to approach gesture analysis. Liddell (1990, 1995, 1998, 2003) and Taub (2001) have shown that the same is true for analysis of signed languages.

On the other hand (to use an appropriate metaphor), gesture analysis is crucial input to mental spaces theory and to cognitive linguistics in general. It is hard to exaggerate the degree to which gesture pervasively embodies the source domain spaces, in representing abstract concepts metaphorically. It is also hard to imagine that the performance of these physical routines in space is as unrelated to the source domain concepts as some have claimed is the case with metaphoric language. Making a fist *is* an instance of solidity, and gesturing forwards *is* an instance of forward motion, unlike the linguistic forms *solid evidence* and *years ahead of us*. Discourse space structure is equally striking: the claim that two compared domains (such as short-term planning memory and chess) are two different mental spaces is enacted in gestural production by assigning them two different physical spaces as loci.

Gesture varies crosslinguistically and crossculturally, as has been shown by McNeill (1992, 2005), McNeill and Duncan (2000), Duncan (2001), Kendon (1990, 1995, 2000, 2004), Haviland (1993, 2000), Núñez and Sweetser (2006) and many others. It is a potentially crucial source of evidence for differences in linguistic cognition between communities. Fascinatingly, the limitations and constraints on gesture may show us cognitive universals – potentially even ones which we have not yet noticed in language. It was already well established (Moore 2000) that in language, deictic metaphors for time are universally back-front ego-referent metaphors; this claim is so far well supported by gestural data (Núñez & Sweetser 2006). But would we have known that aspectual gestures about abstract processes therefore necessarily anchor themselves in an origo at the body – and how universal is this generalization, so clearly manifested in English? What exactly does it mean about cognition?

Gesture, it seems, offers potential evidence relating to linguistically expressed conceptual structure in a huge range of areas of interest to a linguist – concrete linguistic meaning, abstract linguistic meaning, discourse regulation and interaction, indexicality and viewpoint, and doubtless much more. All we have to do is reach out and observe the evidence at hand.

References

Alibali, M. W. (1994). *Processes of cognitive change revealed in gesture and speech.* Ph.D. dissertation, University of Chicago.

Alibali, M. W. & Goldin-Meadow, S. (1993a). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468–523.

Alibali, M. W. & Goldin-Meadow, S. (1993b). Modeling learning using evidence from speech and gesture. *Proceedings of the 15th annual conference of the Cognitive Science Society*, 203–208.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language*. Cambridge, England: Cambridge University Press.

Bavelas, J., Chovil, N., Lowrie D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469–489.

Butterworth, B. & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 347–360). New York: Plenum.

Butterworth, B. & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review, 96*, 168–174.

Calbris, G. (1990). *The semiotics of French gesture*. Bloomington: Indiana University Press.

Cienki, A. (1998a). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (Ed.), *Discourse and Cognition* (pp. 189–204). Stanford CA: CSLI Publications.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H. (1973). Space, Time, Semantics and the Child. In T. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 27–63). New York: Academic Press.

Dancygier, B. (1998). *Conditionals and prediction.* Cambridge: Cambridge University Press.

Dancygier, B. & Sweetser, E. (2005). *Mental spaces in grammar: Conditional constructions.* Cambridge: Cambridge University Press.

Duncan, S. (2001). Perspectives on the co-expressivity of speech and co-speech gestures in three languages. In *Proceedings of the 27th Annual Meeting of the Berkeley Linguistics Society*. Berkeley CA: The Berkeley Linguistics Society.

Dudis, P. G. (2004a). *Depiction of Events in ASL: Conceptual Integration of Temporal Components*. Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Dudis, P. G. (2004b). Body partitioning and real-space blends. *Cognitive Linguistics, 15*(2), 223–238.

Engle, R. A. (2000). *Towards a theory of multimodal communication: Combining speech, gestures, diagrams and demonstrations in instructional explanations*. Ph.D. dissertation, School of Education, Stanford University.

Fauconnier, G. & Turner, M. (1996). Blending as a central process of grammar. In A. Goldberg (Ed.), *Conceptual Structure, Discourse and Language* (pp. 113–130). Stanford: CSLI.

Fauconnier, G. & Turner, M. (2002). *The Way We Think.* Basic Books.

Fillmore, C. J. (1997 [1971]). *Lectures on Deixis*. Stanford CA: CSLI Publications. (originally published by Indiana University Linguistics Club).

Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill (Ed.), *Language and gesture* (pp. 99–117). Cambridge: Cambridge University Press.

Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3/4), 455–479.

Goldin-Meadow, S. (1993). When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents. In K. R. Gibson & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (pp. 63–85). New York, NY: Cambridge University Press.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Science, 3,* 419–429.

Goldin-Meadow, S. (2003). *Hearing gesture*. Cambridge MA: Harvard University Press.

Goodwin, C. (2000). Gesture, aphasia and interaction. In D. McNeill (Ed.), *Language and gesture* (pp. 84–98).

Grady, J. (1997). *Foundations of meaning: Primary metaphors and primary scenes.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Grady, J. (1998). The "Conduit Metaphor" revisited: a reassessment of metaphors for communication. In J.-P. Koenig (Ed.), *Discourse and Cognition: Bridging the Gap* (pp. 205–217). Stanford, CA: CSLI.

Haviland, J. B. (1993). Anchoring, iconicity and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology, 3,* 3–45.

Haviland, J. B. (2000). Pointing, gesture spaces and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13–46). Cambridge: Cambridge University Press.

Hill, C. (1982). Up/down, front/back, left/right: A contrastive study of Hausa and English. In J. Weissenborn & W. Klein (Eds.), Here *and* there: *Crosslinguistic studies on deixis and demonstration* (pp. 11–42). Amsterdam: John Benjamins.

Hopper, P. & Traugott, E. Closs (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Horn, L. (1985). Metalinguistic negation and pragmatic ambiguity. *Language 61*, 121–174.

Horn, L. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge MA: MIT Press.

Johnson, C. (1996). Learnability in the acquisition of multiple senses: SOURCE reconsidered. *Proceedings of the 22nd Annual Meeting of the Berkeley Linguistics Society*, 469–480.

Johnson, Ch. (1999). *Constructional grounding: The role of interpretational overlap in lexical and constructional acquisition.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Kegl, J. & Iwata, G. (1989). Lenguaje de Signos Nicaragüense: A pidgin sheds light on the 'creole' ASL. In R. Carlson et al. (Eds.), *Proceedings of the Fourth Meeting of the Pacific Linguistics Conference* (pp. 266–294). Eugene OR: Dept. of Linguistics, University of Oregon.

Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. De Graff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 179–237). Cambridge, England: Cambridge University Press.

Kendon, A. (1990). Gesticulation, quotable gestures and signs. In M. Moerman & M. Nomura (Eds.), *Culture embodied* (pp. 53–77). Osaka: National Museum of Ethnology.

Kendon, A. (1995). Gestures as illocutionary and discourse markers in Southern Italian conversation. *Journal of Pragmatics, 23*, 247–279.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture and cognition meet.* Mahwah NJ: Lawrence Erlbaum Associates.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M (1999). *Philosophy in the Flesh: The embodied mind and its challenge to Western thought.* New York: Basic Books.

Lakoff, G. & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York: Basic Books.

LeBaron, C. & Streeck, J. (2000). Gestures, knowledge and the world. In D. McNeill (Ed.), *Language and gesture* (pp. 118–138). Cambridge: Cambridge University Press.

Levinson, S. C. (1997). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology, 7*(1), 98–131.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*. Cambridge: Cambridge University Press.

Liddell, S. (1990). Four functions of a locus: Reexamining the structure of space in ASL. In C. Lucas (Ed.), *Sign Language Research: Theoretical Issues* (pp. 176–198). Washington, D.C.: Gallaudet University Press.

Liddell, S. (1995). Real, surrogate and token space: grammatical consequences in ASL. In K. Emmorey & J. Reilly (Eds.), *Language, Gesture and Space* (pp. 19–41). Hillsdale NJ: Lawrence Erlbaum Associates.

Liddell, S. (1998). Grounded blends, gestures, and conceptual shifts. *Cognitive Linguistics, 9*(3), 283–314.

Liddell, S. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.), *Language and gesture* (pp. 331–357). Cambridge: Cambridge University Press.

Liddell, S. (2003). *Grammar, gesture and meaning in American Sign Language.* Cambridge: Cambridge University Press.

Mandel, M. (1977). Iconic devices in American Sign Language. In L. Friedman (Ed.), *On the other hand* (pp. 55–107). London: Academic Press.

McNeill, D. (1992). *Hand and mind.* Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago: University of Chicago Press.

McNeill, D. (Ed.). (2000). *Language and Gesture.* Cambridge, England: Cambridge University Press.

McNeill, D. & Duncan, S. (2000). Growth points in thinking for speaking. In McNeill (Ed.), *Language and Gesture* (pp. 141–161). Cambridge, England: Cambridge University Press.

Mayberry, R. I., Jaques, J., & DeDe, G. (1998). What stuttering reveals about the development of the gesture-speech relationship. In J. Iverson & S. Goldin-Meadow (Eds.), *Nature and Function of Gesture in Children's Communication* (pp. 89–100). San Francisco: Jossey-Bass.

Moore, K. E. (2000). *Spatial experience and temporal metaphors in Wolof: Point of view, conceptual mapping, and linguistic practice.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and gesture* (pp. 186–198). Cambridge: Cambridge University Press.

Núñez, R. & Sweetser, E. (2006). With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science, 30*, 1–49.

Ohala, J. J., Hinton, L., & Nichols, J. (Eds.). (1994). *Sound Symbolism.* Cambridge: Cambridge University Press.

Özyurek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture* (pp. 64–83). Cambridge: Cambridge University Press.

Parrill, F. & Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture, 4*(2), 197–219.

de Ruiter, J. P. (1995). Why do people gesture at the telephone? In M. Bieman & M. Woutersen (Eds.), *Proceedings of the Center for Language Studies Opening Academic Year 1995–96* (pp. 49–56). Nijmegen: Center for Language Studies.

Sapir, E. (1929). A study in experimental symbolism. *Journal of Experimental Psychology, 12,* 225–239.

Sizemore, M. & Sweetser E. (in press). Personal and interpersonal gesture speces: Functional contrasts in language. In A. Tyler, Y. Kim, & M. Takada (Eds.), *Language in the Context of Use: Cognitive and Discourse Approaches to Language and Language Learning*. Berlin.

Slobin, D. I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for Linguistic Relativity* (pp. 107–138). Amsterdam: John Benjamins.

Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking.' In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, England: Cambridge University Press.

Slobin, D. I. (1987). Thinking for speaking. In *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society* (pp. 435–445). Berkeley, CA: The Berkeley Linguistics Society.

Smith, N. J. (2003). Gesture and beyond. Honors thesis, Program in Cognitive Science, UC-Berkeley.

Sweetser, E. (1987). Metaphorical Models of Thought and Speech: a comparison of historical directions and metaphorical mappings in the two domains. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 446–459). Berkeley, CA: The Berkeley Linguistics Society.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure.* Cambridge: Cambridge University Press.

Sweetser, E. (1992). English Metaphors for Language: Motivations, conventions, and creativity. *Poetics Today, 13*(4), 705–724.

Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. In *Actes du 16e Congrès International des Linguistes* (CD ROM). Oxford: Elsevier

Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. In CD-ROM *Proceedings of the 16th International Congress of Linguists*, held in Paris in July 1997.

Sweetser, E. (2001). Deixis, viewpoint and modality. Plenary talk at the International Cognitive Linguistics Conference, July 16–20, 2001 at the University of California at Santa Barbara.

Sweetser, E. (2002). Conceptual blending and viewpoint. Plenary talk at the international workshop "The Way We Think," at Odense University, Demark, August 19–23, 2002.

Sweetser, E. (2003). Literal and metaphorical viewpoint in gesture. Paper presented at the 8th International Cognitive Linguistics Conference, at the University of La Rioja, Spain, July 2003.

Taub, S. F. (2001). *Language from the Body: Iconicity and metaphor in American Sign Language*. Cambridge, England: Cambridge University Press.

Traugott, E. Closs (1982). From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In W. P. Lehmann & Y. Malkiel (Eds.), *Perspectives on historical linguistics* (pp. 245–71). Amsterdam: John Benjamins.

Traugott, E. Closs (1989). On the rise of espistemic meanings in English: An example of subjectification in semantic change. *Language, 65*, 31–55.

Turner, M. & Fauconnier, G. (1995). Conceptual integration and formal expression. *Journal of Metaphor and Symbolic Activity*, *10*(3), 183–204.

Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. R. Ferguson, & E. A. Moravcsik (Eds.), *Universals of Human Language, Vol. 2.* (pp. 527–568). Stanford: Stanford University Press.

Methodology for multimodality

One way of working with speech and gesture data

Irene Mittelberg

1. Introduction

Multimodality, besides entailing at least two modes of expression, also engages at least two modes of interpretation. For instance, when language and gesture coincide, the body of the speaker is as much involved in expressing thoughts, attitudes, or emotions, as are the linguistic articulators. Simultaneously, while listening, the addressee/observer draws on both her sense of audition and vision to make sense of what the interlocutor is conveying. Compared to the various other types of multimodal communication that have come to predominate contemporary media landscapes, this kind of fusion of two modes of expression, speech and gesture, is perhaps the most basic one. Yet given the complexities of human communicative behavior, researchers wishing to investigate its specific aspects face a variety of methodological questions and challenges. These may range from data access to combining methods that do justice to both the semiotic idiosyncrasies of each modality and to the ways in which the different modes interact in a given context.

This chapter provides an example of the empirical steps involved in working with multimodal data consisting of spontaneous speech and its accompanying gestures. It is meant as a companion to the introductory chapter by Mittelberg, Farmer, and Waugh (this volume), which offers a general overview of approaches to discourse and corpus analysis. It is important to realize that the routine presented below is only one of many ways of transcribing, coding, and analyzing co-speech gesture (see also Sweetser this volume). While special attention is paid to gesture-speech co-expression, transcribing naturally occurring speech data, which represents the logical first step in multimodal data transcription, will also be discussed. Overall, the chapter reflects and advocates the trend, noticeable already for some time in various areas of linguistic inquiry, to account for bodily semiotics in the recording and analysis of spoken discourse (e.g., Cienki 1998; Clark 1996; De Stefani fc.; Duranti 1997; Goodwin 1986; Kendon 2004; Lantolf & Thorne 2006; Levinson 2003; McNeill 1992, 2005; Müller 1998, 2004a/b; Schegloff 1984; Streeck 1993, 1994; Sweetser 1998; Tao 1999).

Despite this substantial body of work and the ubiquity of gesture in human communication, manual gestures and other types of bodily semiotics, such as gaze and head movements, have received considerably less scholarly attention than language itself. One of the pragmatic reasons for this is that it took the advent of film and video technology to actually capture ephemeral body movements evolving in space and time. While portraying visual imagery, gestures are not simply visual signs, but dynamic visuo-spatial, or "motor signs" (Jakobson 1987: 474ff.). Spontaneous co-speech gestures are, for the most part at least, polysemous signs that are inseparable from the human body and its physical and sociocultural environments. They thus differ considerably from visual signs that one can contemplate and analyze also after they are taken out of their original context of production. If we, for instance, compare gestures to static sculptures or pictorial signs captured on paper or canvas, we realize that they are extremely fluid figures vanishing as quickly as they take shape. Only with the help of an artificial medium can we truthfully represent and examine co-speech gestures as holistic gestalts of locally-anchored meaning-making.

Having access to advanced video technology has evidently opened up whole new possibilities to the language analyst whose work with usage data had usually been confined to audio recordings. These new ways of looking at language might not be exactly comparable to the revolution ignited by the invention of motion pictures and the altered perception of the world it engendered. Yet the evolution of the film medium has been a precondition for the study of gesture as it is practiced in today's digital era: moving images have made the reproduction of bodily motion in real-time possible, and constitute valuable visual resources. Crucially, video materials facilitate the repeated viewing of the same motion/speech event, even in slow motion and frame-by-frame in frozen bits and pieces, which is crucial for gesture analysis. The opening paragraph of Solnit's book on Eadweard Muybridge, who in the late nineteenth century succeeded in capturing high-speed motion photographically, thus preparing the ground for movies in the U.S., succinctly depicts the tight interrelation between perception, media, art, science, and consciousness:

> In the spring of 1872 a man photographed a horse. The resulting photograph does not survive, but from this first encounter of a camera-bearing man with a fast-moving horse sprang a series of increasingly successful experiments that produced thousands of extant images. The photographs are well known, but they are most significant as the bridge to a new art that would transform the world. By the end of the 1870s, these experiments had led to the photographer's invention of the essentials of motion-picture technology. He had captured aspects of motion whose speed had made them as invisible as the moons of Jupiter before the telescope, and he had found a way to set them back in motion. It was as though he had grasped time itself, made it stand still, and then made it run again, over and over. Time was at his command as it had never been at anyone's before. A new world had opened up for science, art, for entertainment, for consciousness, and an old world had retreated farther. (Solnit 2003: 3)

Working with co-speech gesture provides insights into these dimensions of the new world, especially in terms of temporality and motion through space. Indeed, while closely watching video sequences, the analyst might find herself mesmerized by the poetry arising from the exactly timed interplay between rhythmic imagery and speech prosody (cf. McNeill (2005) and Furuyama (2000b) on the poetic function, Jakobson (1960) in gesture). At the same time, working with video data puts additional demands on the researcher who needs to be both a careful listener and an astute observer. Gesture studies thus do not only put the speaker's body into the picture, they also engage the linguist's eye, behind the camera and in front of the computer screen (cf. Goodwin 2001; Settekorn 1993, 1996, 2003). Investigating linguistic and visual action in this way is not without theoretical repercussions. It lends a certain dynamism to language study which is also reflected in views of language and grammar that bring to light the dynamic side of cognitive and semiotic processes (e.g., Fauconnier & Turner 2002; Jakobson 1987; Jakobson & Pomorska 1980; Hopper 1998; McNeill 2005; Müller 2004b; Peirce 1960; Waugh et al. this volume).

As is the case with approaches to discourse concerning the verbal side of communication, there exists to date no unified empirical method used for gesture analysis. This lack of methodological unity can in part be explained by the fact that gesture researchers come from a variety of disciplines with their own prominent research questions and methodological traditions: among others, psychology, anthropology, linguistics, human development, and communication. From a psycholinguistic perspective, methods of coding and annotating co-speech gesture have been primarily shaped by McNeill (e.g., McNeill 1992, 2000, 2005; McNeill et al. 2001), Duncan (Duncan 2003; McNeill 2005), and Goldin-Meadow (2003) (additional approaches will be discussed below, see also Sweetser this volume; and Mittelberg et al. this volume).[1]

The methodological approach presented here is based on Mittelberg's (2006) study of metaphor and metonymy in gestural representations of grammar. It has particularly been inspired by the methods of transcription, coding, and analysis developed by members of the McNeill Lab at the University of Chicago (McNeill 1992), Cornelia Müller (1998, 2004a) at the Freie Universität Berlin, and Rebecca Webb (1996) at the University of Rochester.[2] The corpus built for the purpose of this research comprises twenty-four hours of naturalistic academic discourse and co-speech gestures produced by four linguists (all native speakers of American English) videotaped during linguistics courses at two major American universities. Given the constraints on space, we cannot reenact the entire research enterprise in great detail, nor can we go much into theoretical motivations or content analysis. While the empirical methods were developed based on both theoretical interests and the nature of the data, the focus here will primarily be on methodological considerations and practical issues of general interest. The intent of the chapter is to convey an idea of some of the decisions one needs to make, along with the different phases of the necessary hands-on work. Along the way, I will point to additional options and more comprehensive sources that might be useful and inspiring to anyone who is considering the option of collecting and exploiting multimodal usage data. Also, the procedure

1. For a concise overview of the field see Kendon (1997); for more comprehensive accounts see Calbris (1990), McNeill (1992, 2000, 2005), Müller (1998), Streeck (1993, 1994), and the recently published monographs by Beattie (2003), Kendon (2004), and Müller (2004b). The interested reader is referred to the journal GESTURE (A. Kendon & C. Müller, Eds.) and to the website of the *International Society for Gesture Studies*: www.gesturestudies.com. For information on the *Berlin Gesture Center* visit www.berlingesturecenter.de.

2. For coding and transcription conventions see McNeill (1992, 2005); an up-to-date gesture coding manual, prepared by S. Duncan, can be found on-line at www.mcneilllab.uchicago.edu

I will walk the reader through here is not to be regarded as completely well-rounded from a methodological point of view. As one often experiences with this kind of work, what works and what doesn't work becomes evident during the process or even after the fact. So, although the methodological approach has been adjusted and refined a number of times, it still represents a work in progress. The reader is invited to evaluate the methods developed for the study discussed below and to determine what appears to be relevant in the context of her or his specific research interests.

The chapter is structured as follows: after addressing some preliminary issues regarding research interests, genre, and data elicitation (Section 2), I will describe procedures of video recording and editing (Sections 3 and 4), discourse transcription (Section 5), and annotation of gesture-speech synchrony (Section 6). Finally, coding parameters used to describe physical gesture features (Section 7) as well as ways of accounting for semantic and pragmatic functions of gestures will be discussed (Section 8).

2. Preliminary considerations: Motivations, subject matter, and genre

Before setting out to collect data, the researcher obviously needs to carefully delimit the domain of inquiry. One source of motivation could be to find non-linguistic evidence for conceptual phenomena one has worked on before, such as metaphors for specific ideas or emotions. This was the case for the present study, which turned out to be a continuation of the author's previous work on linguistic and pictorial metaphors for grammar (Mittelberg 2002). One might also get inspired by existing gesture research: gesture and language development (Goldin-Meadow 2003; McNeill 1992, 2005), second language acquisition and teaching (Gullberg 1998; Lantolf & Thorne 2006; McCafferty 2004; Negueruela et al. 2004), conversation analysis (e.g., De Stefani fc., Goodwin 1986, 2001; Schegloff 1984; Streeck 1993, 1994), and (cross-)cultural investigations (e.g., Efron 1972; Kendon 1995, 2004; McNeill 2005; Müller 1998) are only some of the manifold possibilities. Within these areas, one might consider investigating the different forms and functions of one particular gesture (cf. Müller's (2004a) study of the palm-up open-hand gesture), a gesture family (cf. Kendon 2004), or pointing practices (Fricke 2002; Kita 2003). Other possibilities include micro-analyses of the close relationship between linguistic and gestural co-expression of ideas in an unfolding discourse (Kendon 2000; Kita 2000; McNeill & Duncan 2000; McNeill 2005; McNeill et al. 2001). These decisions will determine not only the environment one needs to explore, but also the kind of genre the data will represent. It will also affect the choice of parameters and the level of detail that will ultimately shape the ensuing procedures of transcription and analysis. One should thus keep in mind that genre and context strongly influence discourse pragmatic factors that in turn are likely to motivate a speaker's linguistic choices and the kinds of gestures that will be made.

A look at previous gesture research attests to the tight interrelation of these factors (subject matter, genres, context, communicative behavior, etc.) and conveys an idea of how they are linked to different methods of eliciting and collecting multimodal data. One can broadly distinguish between naturalistic environments (e.g., authentic talk-in-

interaction), experimental conditions, and quasi-experimental conditions.[3] With respect to the latter method, McNeill and colleagues based a significant portion of their investigations on particular visual stimuli, e.g., films and animated cartoons (such as "Canary Row"; see McNeill 1992, 2000, 2005; McNeill & Levy 1982). Immediately after watching the film, participants are asked to recount the story from memory to a listener. This kind of storytelling technique, based on the same visual stimulus (without much verbal content) for all participants, has several advantages: it can be used with speakers of different age groups, including children, speakers of different languages, speakers with certain impairments (e.g., brain injuries), etc. It also provides a common denominator in terms of visual imagery, semantic content, sequence of events, and narrative structure. This work has especially shed light on iconic gestures that render the speakers' mental images of concrete objects, actions, and settings as they were seen in the cartoon. Being aware of what the participants try to convey in words and gestural imagery, the researcher does not rely exclusively on the speech content to determine the meaning of a gesture. Moreover, she can identify patterns within and across the different renditions of a single scene. When comparing narrations performed in different languages, this sort of data allows one to discern tendencies in the distribution of semantic features across modalities, thus revealing aspects of information management and how gestural action seems to be intertwined with the phrasal organization of the concurrent speech (see also Kendon 2004: 113). For instance, there is interesting work on the ways in which the various semantic aspects of complex motion events (involving path and/or manner) are encoded in speech and/or gesture. These kinds of investigations have yielded compelling insights into typological differences and linguistic relativity (cf. Duncan (2003) on verbal aspect; cf. Kita & Özyürek (2003), Lantolf & Thorne 2006; McNeill (1992, 2000, 2005); McNeill & Duncan (2000); and Müller (1998) on gesture research illuminating aspects of thinking-for-speaking (Slobin 1987, 1996) and different patterns of speech-gesture synchrony in verb-framed vs. satellite-framed languages (Talmy 1985).

Narratives elicited in quasi-experimental settings such as the one described above are an especially well-represented genre in gesture research. Such set-ups allow for the control of environmental factors such as physical setting and participant constellation. Participants are typically asked to sit down in a chair in front of a dark, uni-colored background so that arm and hand movements can be easily discerned when analyzing the video data. As pointed out earlier, there is also a fair amount of work done on different types of conversational data (e.g., Müller 1998, 2004a/b; Kendon 2004; Sweetser this volume; Streeck 1993; Tabensky 2001). Except for an increase in speaker number, the speaker constellation in talk-in-interaction can be similar to the one just described, i.e., in that the interlocutors are seated and do not move around. However, the range of potential research contexts is obviously vast; there is a host of scenarios (two-party or multi-party interaction) in which multimodal communicative practices can be studied (e.g., service encounters, workplace activities, classroom interaction, etc.). Work exploring diverse in-

3. The focus of discussion will be on discourse-based work; for experimental work on gesture see, for instance, Beattie (2003), Goldin-Meadow (2003), Emmorey & Casey (2001), Parrill (2003), and contributions in Emmorey & Reilly (1995) and in McNeill (2000).

teractional settings has shown that the activities in which participants engage shape the interplay between language and bodily communication in a dynamic fashion (cf. contributions in De Stefani fc.; Streeck 2002). Yet different dynamics can be observed in small group discussions centered around situated cognitive activities such as solving math problems (Goldin-Meadow 2003; McNeill 1992; Smith 2003), reviewing architectural models (LeBaron & Streeck 2000), interpreting archeological excavations (Goodwin 2003), giving presentations in science classrooms (Kress et al. 2001; Ochs et al. 1996; Roth 2003), or giving instructions in expert-laymen exchanges (Furuyama 2000a; Haviland 2000; Streeck 2002; Williams 2004).

Expository prose, such as academic talks and lectures about specific topics, is another genre that has received attention from gesture researchers (e.g., Sweetser 1998 this volume; Parrill & Sweetser 2004; Webb 1996) and represents the kind of data on which the present methods discussion is based (see also Mittelberg 2006). When asking the subjects of the study for permission to videotape their lectures (in regards to appropriate protocols in the use of human subjects for data collection, please see Gonzalez-Marquez et al. and Wilcox & Morford, both in this volume), the purpose of the project was framed in very general terms and the fact that gesture was of interest was not revealed so that the speakers would talk and act as naturally as possible. The conditions were thus in no way controlled. The focus was on the teacher giving lectures in linguistics courses, and not on teacher-student interaction, which would certainly be a fruitful next step of inquiry. The reason for choosing this focus came from the realization that for an initial study of metaphoric and metonymic gestural representations of grammar, it would suffice (and already be a complex enterprise) to analyze the teachers' discourse and gestures. Introductory courses seemed to be the right level, as teachers could be expected to make pedagogical efforts when introducing new technical terms, concepts, and theories. Also, the courses were selected such that the data would cover, in addition to general grammatical phenomena, several views of grammar and linguistic theory: generative grammar, emergent grammar (in the context of teaching second languages), and relational grammar. Since linguistic theories are themselves built on specific sets of metaphors, and since gestures are assumed to depict aspects of the source domains of metaphorical mappings, it was hypothesized that the framework talked about in a given instance would influence the kinds of gestures produced to illustrate the speech content.

3. Equipment, physical setting, speaker activities, and videotaping

We will now turn to more practical, hands-on issues involved in video data collection. The following technical equipment was used for audio-video recording: a digital camera (Sony Handycam DCR-TRV900 NTSC), a tripod, and tapes (mini DV cassettes). As the camera has an excellent internal microphone, and the classrooms in which the courses were held were relatively small, no additional microphones were used. However, to achieve superior sound quality of recordings done in relatively big rooms, it is recommended to use wireless microphones and ask the speakers to wear them on the body.

In teaching contexts, speakers usually move about the classroom, write on blackboards, whiteboards, or overhead transparencies, point to information on boards and screens, interact with the audience by turning and/or walking towards students who ask questions, and so forth. These practical issues are not trivial, for when it comes to collecting data, factors such as physical environment, speaker constellation (number of speakers, actions they perform, etc.), and the use of artifacts and space will influence how the video camera needs to be set up in order to ensure that all the physical elements, and especially the speaker's gestures, will be captured as completely as possible. In the case of the study reported here, the video camera was mounted on a tripod placed in the back of the classroom, usually in the middle of the back wall. This choice was motivated by two considerations: first, the full range of potential speaker movements needed to be covered, and second, the investigator tried to be as unintrusive as possible. Once the taping is completed, it is recommended to make back-up copies of each tape and work with the copies, not the originals. Importantly, using digital video technology has the advantage that the data can directly be transferred into the computer to cut and edit the sequences of interest (see Goodwin 1993 for more details and helpful tips regarding video formats, taping procedures, quality of sound and lighting, labeling and storing tapes, software, etc.; see also McNeill et al. 2001 and McNeill 2005 for an example of work with advanced, semi-automated technology).

4. Assessing and editing video data

In a first approach, one might want to view the data several times in order to get an idea of the idiosyncracies of each speaker's linguistic and gestural expression (e.g., is she or he left-handed or right-handed?), the general speech content (in the case of this study: morphology, syntax, etc.), the use of tools in the environment (e.g., overhead, blackboard, chalk, markers, pointers, etc.), and other factors of interest. For purposes of documentation, it is useful to keep a tape content log with information about what each tape contains and other aspects one deems important. Such a log differs from a transcript in that its primary function is to provide a table of contents for each tape, indicating where (using time stamps) specific episodes can be found and whether/why they are of interest. This comes in handy if, in the course of the analysis, one notices particularly interesting things that did not seem to be relevant at first. A content log will also help the researcher to quickly locate specific instances in the entire corpus.

After preliminary data screening, the investigator might choose to do a micro-analysis of a single sequence, or to work with a number of segments representing different speakers, examining the occurrence of one or several gestures across speakers, subject matters, or contexts. Opting for the latter approach, the corpus was assessed from a thematic point of view, selecting and capturing episodes in which gestural representations of grammatical phenomena occurred. Gaze, facial expressions, and movements of the head and torso were not taken into consideration. The goal was to determine how speakers linguistically and gesturally represented particular linguistic units (morphemes, words, phrases, etc.), categories (verb classes, semantic roles, etc.), and structures (clauses, sentences, etc.), syntactic operations (active-passive transformation, subordination, reiteration, etc.), as well as theoretical views of language and grammar more generally.

The software used to edit the video material was the professional editing program FinalCutPro (designed for Macs). After marking the starting and exit time code of each segment, the selected sequences were cut, captured, and saved as separate files (altogether approximately 120 clips), with each clip named according to the speaker and grammatical point talked about in the segment. When editing the clips, one needs to take care not to cut off a speaker in the middle of a sentence and also to capture gestural movements from their onset all the way to their completion, trying to avoid making a cut before the hands fully retract to a rest or neutral position. Working with individual video files has the advantage that copies can be easily made, and that the clips can then be categorized according to speaker, content, gesture type, underlying metaphorical concept, and so forth. In order to facilitate data access for the (often traveling) investigator, advisors, and other gesture researchers interested in viewing the data, all clips (converted to i-movie files) were made available online.

5. Discourse transcription

Transcribing naturally occurring discourse entails many decisions that should ideally, in each case, be guided by the specific research interests and the nature of the data. In essence, the task is to transpose the flow of verbally conveyed information into a visual medium, or record, that one can work on and share with other researchers: "Discourse transcription can be defined as the process of creating a written representation of a speech event as to make it accessible to discourse research" (Du Bois et al. 1993:45). There are various established transcription conventions developed by discourse linguists that may serve as a model (Du Bois et al. 1993, for methods used in the various approaches to discourse analysis see Cameron 2001 and contributions in Atkinson & Heritage 1984; Edwards & Lampert 1993; Jaworski & Coupland 1999). Nonetheless, it is fairly unlikely that one can simply adopt a method that was originally developed for a particular purpose and set of data without making substantial adjustments. In the case of the study discussed here, the outline of discourse transcription provided by Du Bois and colleagues (Du Bois et al. 1993) was adopted and subsequently adjusted. A crucial distinction to make here is to decide what kind of information needs to be represented in addition to the speech content (e.g., pauses, intonation, lengthening, non-vocal noises, etc.) and what kind of information seems to be irrelevant (e.g., detailed information on primary and secondary pitch). The reasons for deploying this particular convention were that the investigator had previously used it to transcribe speech events and that it is widely used, in various adaptations, by conversation analysts and gesture researchers (e.g., the McNeill Lab).

When deciding which transcription method to use, and how to modify it, one of the central questions concerns the degree of detail needed in order to account for what is going on in the data. Generally, one distinguishes between *broad* and *narrow* transcriptions. According to Du Bois et al. (1993:46), *broad* transcriptions usually contain the following types of information: topic of a segment, speaker labels, time stamp of at least the beginning of the segment, and the words spoken (including all truncated words, false starts, self-repair, and vocalizations such as *um, mhm, uh, oh*, etc.). The stream of speech gets decomposed into intonation units (Chafe 1987). When working with conversational data, the speakers' turns, as well as speech overlap, also need to be indicated. Broad transcriptions additionally include intonation contour information (pitch direction, such as rising and/or falling), hesitations, laughter, pauses, truncated words, and uncertain hearings. When doing a *narrow* transcription, the transcriber also includes notations of breathing, accent, prosodic lengthening, tone, and other vocal noises. In view of all the choices one needs to make, it becomes evident that a finished transcript already reveals a lot about the researcher's interests and foci in terms of both theory and analysis (Ochs 1979; see also Waugh et al. this volume).

Whether or not one is interested in investigating bodily communication, it seems worthwhile to exploit the latest recording technologies and videotape talk-in-interaction, or any other kind of communicative event one wishes to analyze. Visual information about the physical setting, environmental surroundings, the speakers' postures, gestures, facial expressions, lip movements, and the object-oriented actions and social interactions participants engage in facilitates the process of transcribing speech considerably. With the help of visual cues, one can get a better grasp not only of what is actually happening in a given speech situation, but also of the details pertaining to the speech delivery. When transcribing spoken discourse solely from audiotapes, the lack of visual input often creates the impression that one does not get it all, and can even mislead the processing and understanding of what is said and referred to (cf. Duranti 1997: 144f.).

Before looking at a sample transcript from the corpus, it should be mentioned that in the present study only the selected speech/gesture segments were transcribed and that all transcriptions were conducted and/or verified by two transcribers. Each clip was first viewed/listened to in its entirety to get an impression of what was talked about and what actions were performed. The next step then was to listen to the sequence very carefully a few times and write down the utterance word by word. This can first be done without much attention to internal structure and prosody. Subsequently, however, it is necessary to identify intonation contours in order to divide the stream of speech into single intonation units (each of which is represented on a separate line as shown below; cf. Chafe 1987, 1998; Du Bois et al. 1993). The transcript shown below represents only the verbal part of a short sequence in which the speaker explains the fact that there are, from an emergent grammar point of view, no *a-priori* grammatical categories; rather, linguistic form is shaped by discourse function. The name of the sequence is indicated in double parentheses.

(1) ((*no a-priori nouns or verbs*))
... *There*fore, _
you *can't* a-priori for instance, _
de*fi=ne*, \
(...) even a *noun* from a *verb*, \
(...) because, \
(..) *verbs* becomin- become nouns,_
(..) and *nouns* can become verbs, /
(..) depending on *how*_
they are *used* in the *dis*course. \

As the procedure of transcribing cannot be described at great length here, I will provide a list with the notational conventions used in the discourse transcripts (adapted from Du Bois et al. 1993; for a fuller account see Du Bois et al. 1992). The information presented below roughly reflects the different steps that go into creating such audio transcripts.

Intonation units:
- each intonation unit appears on a separate line
- truncated intonation units show a double hyphen at the place where the speaker breaks off the intonation unit before completing its projected contour: – -
- truncated words (end of projected word remains unuttered) within intonation units show a single hyphen: -
- transitional continuity of an intonation unit is continuing: ,
- falling terminal pitch movement in an intonation unit: \
- rising terminal pitch movement in an intonation unit: /
- the direction of the terminal pitch movement is level: _

Pauses:
- short unfilled pause: (..)
- medium unfilled pause: (...)
- long unfilled pause indicating duration in minutes and seconds: (....5)
- the sound quality of a filled pause is indicated in round brackets: (*ehm*)

Vocal noises:
- the symbol for laughter is: @ (one symbol for each 'syllable' of the laughter)

Stress and prosodic lengthening:
- a stressed syllable, word or sequence of words are highlighted in *italics*
- preceding segment is lengthened prosodically: "defi=ne"
- prosody was not a major concern here; see the manual prepared by S. Duncan for transcripts where prosody is registered with changes in font size (McNeill 2005: 262, 275ff.)

Unintelligible speech:
- a pair of angle brackets filled with the capital letter 'X': <XXXXX>

It is advised to keep a copy of just the transcribed discourse without any gesture annotation. That way, fine details pertaining to the speech delivery, noticed only later in the process of gesture analysis, can be added. Also, if one later decides to carry out a narrower transcription of a specific episode, one can always go back and add more specific information.

6. Transcribing gesture-speech synchrony

The accurate description of human behavior and social action is admittedly a challenge, and one might wonder whether verbal descriptions are an adequate means of representing visual information and communicative action. Finding a systematic way to document the dynamics of body motion and actions, unfolding in space and time, is already a difficult task. Integrating information on gesture movements with the concurrent speech, prosody and other paralinguistic aspects of talk is even more complex. Gesture researchers have suggested various schemes for how to graphically capture not only the close temporal relationship between speech and co-speech gesture, but also the kinetic features of gestures (e.g., Calbris 1990; Duranti 1997: 144–154; Kendon 2004; McNeill 1992: Chapter 3, appendix; McNeill 2005: 259–286; McNeill et al. 2001; Müller 1998: 175–199, 284ff.; Parrill & Sweetser 2004; Webb 1996). In such comprehensive transcripts, the course of gestural movements (onset, peak, duration, hold, end, etc.) is typically translated into (typo)graphic representations, superimposed onto the written speech transcript (in the form of brackets, different font sizes, highlighting, underlining, glosses, etc.). This process becomes even more intricate if the researcher tries to account for the repeated occurrence of particular gestural features over a longer stretch of discourse, thus identifying gestural patterns of gestural imagery and anaphoric reference (e.g., "growth points" and "catchments" as introduced in McNeill & Duncan 2000; McNeill 2005; McNeill et al. 2001).[4]

Let us now imagine that we are sitting at a computer with a speech transcript in front of us and the corresponding video sequence opened up in the viewer. To arrive at a fine-tuned analysis, we need to view the same clip, and especially chunks of it, over and over again. Using a navigating device (such as Contour ShuttlePro Controller, a tool for video and audio editing) ensures easy going back and forth in slow motion and frame-by-frame, a procedure absolutely necessary to determine where and how exactly a gesture unfolds in relation to the concurrent speech. Before starting the fine-tuned documentation of each gesture, using the already prepared speech transcript, one might want to consider detailing a prose description of the entire sequence. Even though such a description is an optional part of the transcription procedure, it has the merit of requiring the analyst to watch the video data very carefully and to verbalize the global movement of each gesture, how hands interact with artifacts, as well as the ways in which final phases of one gesture may feed into the next one, etc. Capturing these fleeting phenomena in prose has several advantages: prose descriptions complement the obviously extremely condensed information provided by the gesture-annotated transcripts; one can use the descriptions later on to recall and relate the details of a sequence (e.g., when the video material is not handy); and this kind of textual information can build the basis for, and save time regarding, the eventual write-up of the analysis.

The next step in the gesture coding procedure consists of identifying single gestures within a video episode, aiming at accounting for the exact mechanics of speech-gesture synchrony. Due to the fluid nature of gestural movements and their tendency to overlap, it may be difficult at times to determine the boundaries between individual gestural move-

4. In the study discussed here no additional software was used to analyze the speech and body movements. There are, however, various software options that were particularly developed for these purposes; see, for example, the semi-automated methods developed by Francis Quek, provided in McNeill 2005: appendix; McNeill et al. 2001.

ments – that is, to trace a gesture from the moment the articulators (here hands/arms) begin to depart from a position of rest or relaxation until the moment when they return to rest. Kendon (2004: 111) calls such a full "movement excursion" a *gesture unit*. As mentioned above, it requires the repeated slow-motion viewing of chunks of an episode to find the exact moments in speech, i.e., the linguistic segments on which a gesture starts to unfold (the so-called *preparation*), manifests its peak of effort and clearest shape (the *stroke*), may exhibit a sort of hold (the *post-stroke hold*, according to Kita 1993), and then finally comes back to a relaxed position (*retraction* – be it punctual, held, or continued (cf. Müller 1998: 286f. for more variants)). These different movement *phases* make up a *gesture phrase* (Kendon 2004: 110ff.). Importantly, the stroke (if applicable, together with the post-stroke hold) is assumed to express the crucial part of a gesture's meaning, and is the obligatory component of a gesture phrase. Some of the other types of phases mentioned above are optional (for a fuller discussion and transcription examples, see McNeill 1992: 83f., 2005: Appendix; Kendon 2004: Chapter 7). Regarding the question of how exactly the phases of a gesture phrase and speech segments correlate, McNeill (2005: 35) maintains that "[the] *preparation* for the gesture precedes the co-expressive linguistic segment; the *stroke* coincides with this segment; and any pre- and poststroke holds ensure that this speech-stroke synchrony is preserved (Nobe 1996)." As one can imagine, the broad spectrum of gestural actions cannot always be easily accounted for, and one is likely to find hand movements that defy segmentation and classification.

Now, coming back to the issue of annotation, a gesture-phrase may be represented as follows (again, this list does not contain all possible fine-grained dimensions one could account for):

- The linguistic segments that co-occur with a gesture phrase are set off by [square brackets]. Each gesture is assigned an identification number.
- The syllable/word/words on which the **stroke** of the gesture falls is marked in **bold** type face. If not otherwise stated, the preparation phase is understood as the movement leading up to the stroke (i.e., from the onset of the gesture (opening square bracket) until the stroke itself (word(s) marked in bold); the retraction phase is understood as the movement between the end of the stroke phase and the rest position (closing square bracket).
- The duration of a gesture hold (cessation of movement while preserving form and location) is indicated by underlining the co-occurring speech segments. Differences in the execution of hold phases may be captured as follows: solid underline indicates no incorporated movement; dotted underline indicates some movement such as superimposed beats.

The transcript below is an example of an annotated transcript. In this sequence, the speaker demonstrates the word order change implied in transforming an active sentence into a passive sentence by referring to the subject-object inversion as a 'flip-flop.' She produces a gesture starting out with both forearms held vertically and aligned with her shoulders, palms facing her body and fingertips pointing straight up. Then she crosses her arms over her chest (G1, Figure 1). The gestures are numbered, and glosses provide a very brief description of each gesture (abbreviations used below: bh 'both hands,' pcoh 'palms facing center, open hands').

(2) ((*flip-flop passive*))

G1
Arms held vertically, then cross over chest
[The passive basically **flip**flops _

G1
crossing-gesture being held
the subject and object of the sentence, \
a=nd _

G1 still being held
(...) what we find out by forming this particular passive], _

G2
bh, pcoh-box, being held, hands move up and down
is [that the **string** 'John's **sis**ter' **for**ms a cons**ti**tuent, \

G2
box-gesture still being held, move up and down
namely the object of the **verb**, \

G2
still being held, move up and down hands retract to fist
(..) and that's an object **noun** phrase] in fact. \

Following the stroke of the first gesture (G1), which coincides with the mention of the verb "flipflops," the arms-crossed gesture is being held while the speaker keeps explaining the syntactic operation involved (the underline indicates the duration of the gesture hold). She turns her torso leftward towards the overhead screen behind her, walks briefly towards the screen, then turns back and finally faces the audience again. Subsequently, the speaker's arms open up again and merge right into the next gesture (G2): both open hands being held more than shoulder width apart and facing each other, seemingly holding an (imaginary) object, which here represents "a constituent." There are several dimensions of meaning interacting in this gesture: the physical action represents, iconically, the switching around of two things. At the same time, it stands, metaphorically, for the abstract action of 'flip-flopping' two elements in a sentence. This is an example of a frequently found gesture reflecting the metaphorical concept IDEAS ARE OBJECTS (Lakoff & Johnson 1980, 1999). In the data, variants of this gesture were found to refer to linguistic entities of different complexity (categories, words, phrases, sentences, discourses, etc.) in terms of bounded spaces.

7. Physical gesture features

After focusing on aspects of gesture-speech synchrony and the internal structure of gesture phrases, we will now turn to the task of documenting the physical features of gestural communication. In the gesture literature, the most widely used coding parameters are *hand presence* (left and/or right hand), *hand shape, palm orientation, movement* (trajectory and manner), and the *location* in gesture space where a gesture is performed. Throughout the multi-layered process of gesture annotation and analysis, there is a danger of mixing the domain of form, i.e., what one actually observes with the naked eye, and the domain of meaning, i.e., the semantic and pragmatic functions a particular combination of physical features may assume in a given moment. Taking the material side of the semiotic processes as the point of departure meant, in the case of the present study, moving from concrete forms (gesture) to abstract concepts and structures (grammatical phenomena). Below, I briefly discuss how each parameter was dealt with in the present study, and point to additional work that might be of special interests to cognitive linguists.

7.1 Hand shape and palm orientation

In order to categorize and represent *hand shapes*, a data-driven typology of manual signs was developed. Another possibility would be to adopt the form inventory of a signed language such as American Sign Language (ASL) (cf. McNeill 1992: 86–88; Webb 1996). First, the data were searched for hand shapes and arm configurations recurring across speakers and contexts, and then each identified form was assigned a name. For example, one of the most frequently used hand shapes in the data is an open hand with the palm turned upwards, thus building a sort of surface. Here it seemed worthwhile to build on conventions introduced by Müller (2004a) in her study of forms and functions of the palm-up open hand gesture (or, "puoh").[5] Each variant of the open hand gesture was given an abbreviation such as "puoh" indicating the *orientation of the palm*, a short name evoking the degree of openness of the hand ("tray," "cup," etc.), as well as an indication of which hand performed the gesture. So for instance, "puoh-tray-lh" stands for a flat palm-up open hand, evoking the shape of a tray, produced with the left hand.

Other gestures were observed that were simultaneously performed with two hands, thus exhibiting an internal structure (cf. Kendon 2004: 104, 275ff.). For example, "pcoh-box-bh" stands for another frequently observed gesture consisting of two hands held apart, with both palms being held vertically and facing each other as if they were holding an object (such as a shoe box) between them (i.e., "pcoh" stands for "palm-center open hand," with "center" denoting the direction that the palm is facing). Numerous other hand shapes were identified which cannot be discussed here (for details see Mittelberg 2006).

7.2 Location in gesture space

Space is the natural habitat of gestures. Speakers may exploit the space around them to depict the location of objects, people, places, events, and concepts, as well as the spatial relationships among things, a task usually more difficult to master with purely linguistic means. The range, organization, and preferred use of a person's *gesture space* is condi-

5. In the literature, this kind of open hand gesture is also referred to as "palm presentation" gesture (Kendon 2004: 264) or "a gestural version of the conduit metaphor" (McNeill 2005: 45; cf.: Parrill 2003).

tioned by factors such as age, gender, cultural background, and personal style (cf. Calbris 1990; Emmorey & Casey 2001; Goldin-Meadow 2003; Kendon 1997, 2004; McNeill 1992, 2005; Müller 1998). Also, gesture space is relative to, and constituted by, the position and posture of the speaker-gesturer who, in each communicative act, sets up the coordinates of gesture space around her. Here, several factors may come into play: the dimensions and movements of the speaker's body and gestural articulators, artifacts and physical setting, and, if applicable, also the interpersonal space spanning between herself and her interlocutor(s) (see McNeill (2005: 159ff.) on how participants may (re)shape shared gesture space).

The *location* of a gesture can be described from various angles: relative to the gesturer's body, relative to previously or subsequently produced gestures, or relative to the addressee's gesture space. One of the central interests of the present study was the spatialization of abstract information, i.e., ways in which the speakers' use of the different dimensions of gesture space could be linked to spatial metaphor, and, in particular, to particular theoretical models of grammar. This is where different diagrammatic representations of linguistic structure come to light. For example, horizontal lines running from the left to the right of the speaker may depict a linear sentence model; tree diagrams (generative grammar), by contrast, exploit horizontal, vertical, and diagonal axes, and so forth. Such gestural diagrams appear to serve as virtual grids in gesture space providing slots where imaginary objects, standing in for abstract grammatical units, can be placed and subsequently referred to (cf. Mittelberg fc./a on different types of diagrammatic iconicity in gesture (Peirce 1955)).

It should be mentioned that for the purpose of documenting the exact locations where gestures occur, gesture researchers have developed virtual grids dividing the space around the speaker into compartmentalized regions. The system McNeill (1992, 2005) has proposed shows a shallow disc consisting of concentric squares superimposed on a drawing of a seated person (reflecting the experimental set-up described earlier). With the help of such a system the density of occurrence of gestures in specific sectors of gesture space can be represented, thus revealing where certain gesture types tend to occur, e.g., close to certain body parts (head, chest, etc.), in the center versus the periphery of gesture space, on upper vs. lower planes, etc. To give an example, McNeill (1992: 88) identified a correlation between gesture type and location of occurrence. In cartoon narrations by American university students, iconic gestures (depicting concrete objects) were predominately performed in the immediate center, metaphoric gestures (depicting abstract entities) in the lower center, and pointing gestures in the periphery (for definitions of the different gesture categories used for the study cf. McNeill (2005: 38ff.)). In a further step, these findings were supplemented by a frequency analysis of gesture types in correlation with different discourse contexts: narrative clauses (promoting the development of the story, i.e., the plot line) versus extranarrative clauses (descriptions of the setting and characters, summaries, comments, etc.). McNeill (1992: 92ff.) found that whereas iconic gestures and abstract pointing gestures occur primarily in narrative clauses, metaphoric gestures appear chiefly in extranarrative clauses (for details regarding the relationship between discourse demands and saliency of certain gesture dimensions see McNeill 2005: 42ff.).

From a cognitive linguistics perspective, the relationship between space, embodied cognition, and gesture is particularly fascinating and constitutes a promising domain of inquiry. The interested reader is referred to work on dynamic representations of spatial concepts, spatial arrays, frames of reference, spatial portrayals of time, and mental spaces (e.g., Haviland 2000; Levinson 1997, 2003; Nuñez & Sweetser 2006; Sweetser this volume). Additional insights may be gained from consulting work on the logic and use of space in signed languages (e.g., Dudis 2004; Emmorey & Reilly 1995; Liddell 2003; Milford & Wilcox this volume; Taub 2001).

7.3 Movement

Gestures typically involve some *movement* through space. As opposed to the (unmarked) flux of hand movements, a gesture hold can be regarded as a marked instance (Waugh 1982). There are various types of gestural movements. For example, a movement of a hand can result in the evocation of a form (such as the size and shape of a table), it may be influenced by the object involved in the action that gets imitated by the hand movement (such as writing something down with a pen). A gesture may also simply imitate a manual action (such as waving at somebody) or the manner and/or speed of a movement executed by a person or an object. The hand movements observed in the present data also exhibit several intrinsic logics: movements evoking linear traces (along horizontal, vertical, or diagonal axes) or nonlinear traces (imitating the shape of a wave, circle, or arch); pointing gestures whose direction and range depend on the location of the object or person pointed at; object-oriented actions such as placing something; and genuine motor actions such as two hands rotating around each other. In keeping with the notational conventions used for hand shapes, the prominent movement patterns observed in the data were given labels that inform about its trajectory and manner. For example, "hori-trace-rh" signifies a horizontal line traced in the air with the right hand, and "wrist-rota-lh" refers to a wrist rotation performed with the left hand.

As a result of the steps described above, a set of image-schematic and motor patterns emerged from the data, some of which evoke geometric figures (circles, triangles, squares, etc.) and/or image schemas proposed in the literature (Johnson 1987; Lakoff 1987; Mandler 1996; see also contributions in Hampe 2005, especially Cienki's experimental study on image schemas and gesture). Examples of image schemas evoked in the gesture modality include support, containment, source-path-goal, balance, scale, iteration, front-back, part-whole, link, and forced motion (see Mittelberg fc./b). The next task was to find instantiations of each of the identified shapes and movement patterns across speakers. This was not possible for all categories, but it was possible for most and thus allowed the researcher to discern semantic and pragmatic commonalities and differences, taking into consideration the concurrent speech.

8. Semantic and pragmatic functions of gestures

Depending on one's research interests and theoretical approach, the lens one decides to use to detect meaning in gestures will naturally influence what one actually sees. After attending to the material side of gestures, the eye of the linguist finally becomes engaged in the task of visual interpretation which is, compared to our natural way of perceiving fluid gestural movements, aided and intensified by the repeated viewing of segments. Watching gestures in slow motion and frame-by-frame reveals nuances of bi-modally achieved meaning-making one might otherwise miss. Gestures convey not necessarily the same semantic aspects as the speech segments with which they coincide. It has been shown that spontaneous gestures may reinforce or complement what is expressed verbally in various ways (cf. Kendon 2000, 2004; McNeill & Duncan 2000; McNeill 2005).

Whether one chooses to adopt an existing gesture classification system or decides to design one's own, it is important to realize that in a single act of gestural signification, as in most sign processes, multiple semiotic modes tend to interact (i.e., iconicity, indexicality, symbolization, according to Peirce 1995, 1960). Many gesture researchers have realized that multifunctional signs such as co-speech gestures hardly fit into discrete categories. The gesture categories originally proposed by McNeill (1992) are thus not to be understood as mutually exclusive (i.e., *beats, iconics, metaphorics,* and *emblems*). As a matter of fact, McNeill (2005:41) has recently emphasized that in one and the same gesture different semiotic and functional dimensions may layer to various degrees and that speaking of categories may be misleading: "We should speak instead of *dimensions* and say *iconicity, metaphoricity, deixis*, '*temporal highlighting*' (beats), *social interactivity*, or some other equally unmellifluous (but accurate) terms conveying dimensionality" (italics in the original; for examples and overviews of established paradigms and latest developments cf. Calbris 1990; Efron 1972; Kendon 2004; McNeill 1992, 2000, 2005; Müller 1998, 2004a/b; Parrill & Sweetser 2004; inter alia).

In her functional gesture typology, Müller (1998) distinguishes gestures whose primary function is "referential" (i.e., depicting objects, attributes, actions, etc.) from those that are rather "discursive" (pertaining to the structure of the discourse), or "performative" (i.e., gestures that have a similar function as speech acts). Depending on the specific nature of the referent of the gestural sign, Müller further distinguishes referential gestures that represent concrete entities from those that represent abstract entities (see also Cienki 2005; and Müller 2004a/b). Müller also stresses the point that one and the same gesture may fulfill several functions at the same time (see also Kendon 2004).

The issue of layering semiotic and functional modes comes clearly to the fore when one investigates abstract subject matters and thus enters the domain of metaphor. Metaphoric gestures, referring to abstract concepts and relations, are said to be iconic in the sense that they may depict aspects of the concrete source domain of the underlying metaphorical mapping. As demonstrated above, abstract grammatical categories such as a "constituent" may be gesturally depicted in the form of a manipulatable object, such as an imaginary box seemingly held between the speaker's hands. Although the speech in this example is not metaphorical but technical in nature, the image schema OBJECT surfaces in this manual constellation, hence reflecting the metaphorical concept IDEAS ARE OBJECTS or CATEGORIES ARE CONTAINERS (Lakoff & Johnson 1980, 1999).[6] Proposing a dynamic metaphor theory, Müller (2004b) suggests that gestures may also awaken

6. To represent this conceptual-semiotic linkage in the gesture-annotated transcript, one can, for instance, add two separate lines below the speech line: one describing the iconic (and/or deictic) qualities

metaphors that are supposed to be dead, or inactive in our conceptual system. In other words, while a linguistic expression based on an inactive metaphor might be perceived as literal, the metaphorical understanding still materializes in the accompanying gesture (for an overview of recent advances in research on metaphor and gesture see contributions in Cienki & Müller fc.).

To account for the specific meaning-making processes in the data of the study discussed above, it was necessary to develop a theoretical framework combining several different strands: traditional semiotic theories (Jakobson 1956, 1964, 1987; Peirce 1955, 1960), gesture research (e.g., Bouvet 1997, 2001; Calbris 2003; Cienki 1998; Fricke 2004; Kendon 2004; McNeill 1992, 2000; Müller 1998, 2004a; Sweetser 1998; Webb 1996), and cognitive metaphor and metonymy (e.g., Barcelona 2000; Dirven & Pörings 2002; Gibbs 1994; Goosens 1995; Johnson 1987; Lakoff & Johnson 1980, 1999; Lakoff 1987, 1993; Langacker 1993; Panther & Thornburg 2004; Radden & Kövecses 1999; Sweetser 1990; Taub 2001; Wilcox 2004). Central to this approach is Peirce's (1955) view of the sign as a dynamic process in which the different referential modes layer to various degrees. It further is based on the assumption that in any sign-object relationship one of the interacting modes is predominant, thus causing the receiving mind to interpret it as predominantly iconic, indexical, or symbolic (Peirce 1960), presupposing a relative hierarchy of the various modes (cf. Jakobson 1964, 1966, 1974). Furthermore, the notion of salience played a role in determining those qualities of a gestural gestalt that contribute most significantly to its meaning and function. For example, in certain cases, the movement proves to be more salient with respect to the meaning of a gesture than the particular shape of the hand performing the movement (e.g., in certain gestures that trace the dimension of a sentence in the air, from left to right of the speaker, whereby it does not matter whether the hand tracing is a relaxed flat hand or whether the fingers are configurated in a particular way); in other cases, the hand shape is more salient than the contextual movements (e.g., in the case of the emblematic ring gesture, also known as the OK-sign); and in yet other cases, both dimensions are significant (e.g., a push performed with an open flat palm facing the addressee, thus building a barrier and evoking the idea of 'stop' or 'rejection'). Based on the set of geometric and image-schematic patterns that emerged from the data, the next step then was to examine how the minimal information provided by the hand shapes and movements played into the iconic, metaphoric, and metonymic meaning construction in the ongoing discourse (cf. Mittelberg 2006, fc.).

In light of the different perspectives gesture researchers have been exploring to enhance our understanding of situated practices of language use and its cognitive and sociocultural underpinnings, it becomes evident that gestures may provide valuable non-verbal evidence for, among other things, spatial cognition and the interplay between iconicity, metaphor, and metonymy (see also Sweetser this volume).

---

of the gesture and one indicating what the pictorial content stands for metaphorically (for transcript examples see McNeill 1992: 94ff.; Sweetser 1998; Taub 2001).

9. Concluding remarks

There are many more things to be said and considered here, but I hope that this chapter has provided the reader with a picture of the kinds of decisions and methodological steps that shape empirical work with multimodal usage data. As has become evident, there are diverse approaches and methods to choose from, and the investigator will have to see which ones appear most suitable for a specific research project and also how they might need to be adjusted to exploit as effectively as possible the particular kind of data with which she or he wishes to work.

While the focus here has been on transcribing and describing the material side of multimodal communication, and not on content analysis or theoretical issues, the body of research pointed to throughout the chapter suggests that investigations into the gesture modality may offer a window on the bodily foundation of meaning (see also Sweetser this volume). Working with co-speech gesture data allows the researcher to witness, even if it is just for a moment, how embodied structures become visible in the form of metaphors in speakers' hands, diagrams drawn in the air, or pointing gestures drawing attention to imaginary events being set up in gesture space. Essentially, co-speech gestures are *dynamic* semiotic acts. Compared to, for example, static pictorial metaphors in cartoons and advertisement, metaphorical gestures are strikingly spontaneous and may convey figurative understandings in a comparatively unreflective fashion, even if the concurrent speech is perceived as literal (Müller 2004b).

Overall, it seems that multimodal utterances can only be studied empirically: gestures are inseparable from the speaker's body, and the speaker's body and its communicative practices are always anchored in the physical, interpersonal, and cultural context of the speech event (Kendon 2004). So, no matter how closely we look at instances of bodily semiotics, we may not be able to distinguish the gesturer from the gesture, or to "tell the dancer from the dance" (Kramsch 2002: 1).

Acknowledgments

I would like to thank Linda Waugh, Thomas Farmer, Adam Kendon and Elwys De Stefani for valuable comments on earlier versions of this paper and Cornelia Müller and Rebecca Webb for generously sharing their experience with gesture coding and transcription. I am grateful to Richard Feldman and Andrew Page from the Cornell Language Resource Center for technical support and producing video materials. I thank Allegra Giovine, Joel Ossher, and Daniel Steinberg for their input regarding the methodology presented here and for helping with gesture annotation and analysis. I also thank the Cornell Cognitive Studies Program for financial support.

References

Aitkinson, H. M. & Heritage, J. (Eds.). (1984). *Structures as social action: Studies in conversation analysis*. Cambridge: Cambridge University Press.

Barcelona, A. (Ed.). (2000). *Metaphor and metonymy at the crossroads: A cognitive perspective*. Berlin/New York: Mouton de Gruyter.

Beattie, G. (2003). *Visible thought: A new psychology of body language*. Hove: Routledge.

Bouvet, D. (1997). *Le corps et la métaphore dans les langues gestuelles: A la recherche des modes de production des signes*. Paris: L'Harmattan.

Bouvet, D. (2001). *La dimension corporelle de la parole. Les marques posturo-mimo-gestuelles de la parole, leurs aspects métonymiques et métaphoriques, et leur rôle au cours d'un récit.* Paris: Peeters.

Calbris, G. (1990). *The semiotics of French gestures*. Bloomington: University of Indiana Press.

Calbris, G. (2003). From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture, 3*(1), 19–46.

Cameron, D. (2001). *Working with spoken discourse*. London/Thousand Oaks: Sage.

Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam/Philadelphia: John Benjamins.

Chafe, W. (1998). Language and the flow of thought. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure, volume 1* (pp. 93–111). Mahwah, NJ: Erlbaum.

Cienki, A. (1998). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 189–204). Stanford: CSLI Publications.

Cienki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), in cooperation with J. Grady, *From perception to meaning: Image schemas in Cognitive Linguistics* (pp. 421–442). Berlin/New York: Mouton de Gruyter.

Cienki, A. & Müller, C. (Eds.). (Fc.). *Metaphor and gesture*. Amsterdam/ New York: John Benjamins.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

De Stefani, E. (Ed.). (Fc.). Les données vidéo dans la recherche linguistique Looking at language. [Video data in linguistic research. /Regards sur la langue.] Special issue of the *Bulletin Suisse de linguistique appliquée VALS/ASLA*.

Dirven, R. & Pörings, R. (2002). *Metaphor and metonymy in comparison and contrast*. Berlin/New York: Mouton de Gruyter.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1992). Discourse transcription. *Santa Barbara Papers in Linguistics, 4*, 1–225.

Du Bois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–87). Mahwah, NJ: Lawrence Erlbaum Assoc.

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics, 15*(2), 223–238.

Duncan, S. (2003). Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture, 2*(2), 183–206.

Duranti, A. (1997). *Linguistic anthropology.* Cambridge: Cambridge University Press.

Edwards, J. A. & Lampert, M. D. (Eds.). (1993). *Talking data: Transcription and coding in discourse research.* Hillsdale, NJ: Lawrence Erlbaum.

Efron, D. (1972). *Gesture, race, and culture.* Preface by Paul Ekman. The Hague: Mouton and Co.

Emmorey, K. & Casey, S. (2001). Gesture, thought, and spatial language. *Gesture, 1*(1), 35–50.

Emmorey, K. & Reilly, J. S. (Ed.). (1995). *Language, gesture, and space*. Hillsdale, NJ: Erlbaum.

Fauconnier, G. & Turner, M. (2002). *The way we think: Conceptual integration and the mind's hidden complexities*. New York: Basis Books.

Fricke, E. (2002). *Origo*, pointing, and speech: The impact of co-speech gestures on linguistic deixis theory. *Gesture, 2*(2), 207–226.

Fricke, E. (2004). Origo, Geste und Raum – Lokaldeixis im Deutschen. Doctoral Dissertation, Technische Universität Berlin.

Furuyama, N. (2000a). Gestural interaction between the instructor and the learner in origami instructions. In D. McNeill (Ed.), *Language and gesture* (pp. 99–117). Cambridge: Cambridge University Press.

Furuyama, N. (2000b). The semiotic making of speech and gesture and how it contributes to the structuring of discourse. *Cognitive Studies, 7*(1), 71–77.

Gibbs, R. W., Jr. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge, UK: Cambridge University Press.

Goldin-Meadow, S. (2003). *Hearing gesture: How hands help us think*. Cambridge, MA/London: The Belknap Press of Harvard University Press.

Goodwin, C. (1986). Gesture as a resource for the organization of mutual orientation. *Semiotica, 62*(1–2), 29–49.

Goodwin, C. (1993). Recording human interaction in natural settings. *Pragmatics*, *3*(2), 181–209.

Goodwin, C. (2001). Practices of seeing visual analysis: An ethnomethodological approach. In T. van Leeuwen & C. Jewitt (Eds.), *Handbook of Visual Analysis* (pp. 157–182). London: Sage Publications.

Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where Language, Culture, and Cognition Meet* (pp. 217–241). Mahwah, NJ: Lawrence Erlbaum Associates.

Goossens, L. (1995). Metaphtonymy: The interaction of metaphor and metonymy in figurative expressions for linguistic action. In L. Goosens, P. Pauwels, B. Rudzka-Ostyn, A.-M. Simon-Vandenbergen, & J. Vanparys (Eds.), *By word of mouth. Metaphor, Metonymy and Linguistic action in cognitive perspective* (pp. 159–174). Amsterdam/Philadelphia: John Benjamins.

Gullberg, M. (1998). *Gesture as a communication strategy in second language discourse. A study of learners of French and Swedish*. Lund University Press.

Hampe, B. (Ed.). (2005). *From perception to meaning: Image schemas in cognitive linguistics.* Berlin: Mouton de Gruyter.

Haviland, J. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13–46). Cambridge: Cambridge University Press.

Hopper, P. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure*. Mahwah, NJ: Lawrence Erlbaum.

Jakobson, R. (1990 [1956]). Two aspects of language and two types of aphasic disturbances. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 115–133). Cambridge/London: Harvard University Press.

Jakobson, R. (1960). Closing statement: Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in Language* (pp. 350–377). Cambridge, MA: MIT Press.

Jakobson, R. (1971[1964]). Visual and auditory Signs. Phonetica XI, 216–220; reprinted in *Selected Writings II* (pp. 334–337). The Hague: Mouton.

Jakobson, R. (1990 [1966]). Quests for the essence of language. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 407–421). Cambridge/ London: Harvard University Press.

Jakobson, R. (1990 [1974]). Linguistics in relation to other sciences. In L. R. Waugh & M. Monville-Burston (Eds.), *Roman Jakobson, On Language* (pp. 451–488). Cambridge/ London: Harvard University Press.

Jakobson, R. (1987). On the Relation between Auditory and Visual Signs. In K. Pomorska & S. Rudy (Eds.), *Roman Jakobson, Language in literature* (pp. 467–473). Cambridge/London: Belknap Press of Harvard University Press.

Jakobson, R. & Pomorska, K. (1980). *Dialogues*. Paris: Flammarion. Eng. Trans. 1983. Cambridge, MA: MIT Press.

Jaworski, A. & Coupland, N. (Eds.). (1999). *The discourse reader*. London/New York: Routledge.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason.* Chicago: University of Chicago Press.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *Journal of Pragmatics, 23*(3), 247–279.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology, 26*, 109–128.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (1993). Language and thought interface: A study of spontaneous gestures and Japanese mimetics. Doctoral Dissertation, University of Chicago.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.

Kita, S. (Ed.). (2003). *Pointing: Where language, culture, and cognition meet*. Mahwah, NJ: Lawrence Erlbaum Assoc.

Kita, S. & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 48*, 16–32.

Kramsch, C. (2002). 'How can we tell the dancer form the dance?' Introduction to C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 1–30). London/New York: Continuum.

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: The rhetorics of the science classroom*. London/New York: Continuum.

Lakoff, G. (1987). *Women, fire, and dangerous Things. What categories reveal about the mind.* Chicago: University of Chicago Press.

Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and thought* (pp. 202–251) (2nd ed.). Cambridge: Cambridge University Press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought.* New York: Basic Books.

Langacker, R. W. (1993). Reference-point constructions. *Cognitive Linguistics, 4,* 1–38.

Lantolf, J. P. & Thorne, S. (Ed.). (2006). *Sociocultural theory and the genesis of second language development*. Oxford: Oxford University Press.

LeBaron, C. & Streeck, J. (2000). Gestures, knowledge, and the world. In D. McNeill (Ed.), *Language and gesture* (pp. 118–138). Cambridge: Cambridge University Press.

Levinson, J. (1997) Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology, 7*(1), 93–131.

Levinson, J. (2003) *Space in language and cognition: Explorations in cognitive diversity.* Cambridge: Cambridge University Press.

Liddell, S. (2003). *Grammar, gesture, and meaning in American Sign Language.* Cambridge: Cambridge University Press.

Mandler, J. (1996). Preverbal representation and language. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 365–384). Cambridge, MA: MIT Press.

McCafferty, S. G. (2004). Space for cognition: Gesture and second language learning. *International Journal for Applied Linguistics, 14*(1), 148–165.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: Chicago University Press.

McNeill, D. (Ed.). (2000). *Language and gesture*. Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: Chicago University Press.

McNeill, D. & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

McNeill, D. & Levy, E. T. (1982). Conceptual representations in language activity and gesture. In R. Jarvella & W. Klein (Eds.), *Speech, place, and action* (pp. 271–295). Chichester, England: Wiley & Sons.

McNeill, D., Quek, F., McCullough, K.-E., Duncan, S. D., Furuyama, N., Bryll, R., Ma, X.-F., & Ansari, R. (2001). Catchments, prosody and discourse. *Gesture*, *1*(1), 9–33.

Mittelberg, I. (2002). The visual memory of grammar: Iconographical and metaphorical insights. *Metaphorik.de*, *2/2002*, 69–89.

Mittelberg, I. (2006). Metaphor and metonymy in language and gesture: Discourse evidence for multimodal models of grammar. Doctoral Dissertation, Cornell University.

Mittelberg, I. (fc./a). Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In A. Cienki & C. Müller (Eds.), *Metaphor and gesture*. Amsterdam/New York: John Benjamins.

Mittelberg, I. (fc./b). Geometric patterns in gesture space. In V. Evans & P. Chilton (Eds.), *Language, cognition, and space: The state of the art and new directions.* London: Equinox.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich.* Berlin: Berlin Verlag A. Spitz.

Müller, C. (2004a). Forms and uses of the palm up open hand: A case of a gesture family? In R. Posner & C. Müller (Eds.), *The semantics and pragmatics of everyday gestures. The Berlin conference* (pp. 233–256). Berlin: Weidler Verlag.

Müller, C. (2004b). *Metaphors. Dead and alive, sleeping and waking. A cognitive approach to metaphors in language use*. Habilschrift, Freie Universität Berlin.

Negueruela, E., Lantolf, J. P., Jordan S. R., & Gelabert, J. (2004). The "private function of gesture in second language speaking activity: A study of motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics, 14*(1), 113–147.

Nobe, S. (1996). *Representational gestures, cognitive rhythms, and acoustic aspects of speech: A network/ threshold model of gesture production*. Doctoral Dissertation, University of Chicago.

Núñez, R. E. & Sweetser, E. E. (2006). Aymara, where the future is behind you: Convergent evidence from language and gesture in the cross-linguistic comparison of spatial construals of time. *Cognitive Science, 30*, 1–49.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schiefflin (Eds.), *Developmental pragmatics*. New York: Academic Press. Reprinted in A. Jaworski & N. Coupland (1999) (Eds.), *The Discourse reader* (pp. 167–182). London/New York: Routledge.

Ochs, E., Gonzalez, P., & Jacoby, S. (1996). "When I come down I'm in the domain state": Grammar and graphic representation in the interpretive activity of physics. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 329–369). Cambridge: Cambridge University Press.

Panther, K.-U. & Thornburg, L. L. (2004). The role of conceptual metonymy in meaning construction. *Metaphorik.de, 06/2004*, 91–113.

Parrill, F. (2003). Intuitions and violations of good form in metaphoric conduit gestures. Paper presented at the theme session "Metaphor and Gesture," 8th International Cognitive Linguistics Conference, Logroño, Spain, July 23.

Parrill, F. & Sweetser, E. E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, *4*(2), 197–219.

Peirce, C. S. (1955). Logic as semiotic: The theory of signs (1893–1920). In J. Bucher (Ed.), *Philosophical writings of Peirce* (pp. 98–119). New York: Dover.

Peirce, C. S. (1960). *Collected papers of Charles Sanders Peirce (1931–1958). Vol. I.: Principles of philosophy, Vol. II: Elements of logic.* C. Hartshorne & P. Weiss (Eds.). Cambridge: The Belknap Press of Harvard University Press.

Radden, G. & Kövecses, Z. (1999). Towards a theory of metonymy. In K.-U. Panther & G. Radden (Eds.), *Metonymy in language and thought* (pp. 17–59). Amsterdam/ Philadelphia: John Benjamins.

Roth, W.-A. (2003). From epistemic (ergotic) actions to scientific discourse: The bridging function of gestures. *Pragmatics & Cognition, 11*(1), 141–170.

Schegloff, E. A. (1984). On some gestures' relation to talk. In J. M. Atkinson & J. Heritage (Ed.), *Structures of social action* (pp. 266–296). Cambridge: Cambridge University Press.

Settekorn, W. (1993). Das Auge und der Körper des Linguisten. Für eine Medien- Technik- und Sozialgeschichte (nicht nur) der Sprachwissenschaft, *Linguistik und Literaturwissenschaft, 90/91*, 195–234.

Settekorn, W. (1996). Vor-Film. Zur Geschichte der Inkorporation von Sprache, Bild und Raum. In H. Segeberg (Ed.), *Die Mobilisierung des Sehens. Zur Vor- und Frühgeschichte des Films in Literatur und Kunst* (pp. 13–43). München: Fink.

Settekorn, W. (2003). Representation: The influence of history, culture, and (mass) media on the evolution of a concept central to cognitive science. Presentation given at the EMCL Workshop at Cornell University.

Slobin, D. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistic Society* (pp. 435–445). Berkeley: Berkeley Linguistics Society.

Slobin, D. (1996). From 'thought and language' to 'thinking for speaking'. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 97–114). Cambridge: Cambridge University Press.

Smith, N. (2003). Gesture and beyond. Honors Thesis, Program in Cognitive Science, University of California at Berkeley.

Solnit, R. (2003). *River of shadows: Eadweard Muybridge and the technological Wild West*. London/New York: Penguin.

Streeck, J. (1993). Gesture as communication I: Its coordination with gaze and speech. *Communication Monographs, 60,* 275–299.

Streeck, J. (1994). Gesture as communication II: The audience as co-author. *Research on Language and Social Interaction, 3,* 239–276.

Streeck, J. (2002). A body and its gestures. *Gesture, 2*(1), 19–44.

Sweetser, E. E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press.

Sweetser, E. E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. *Actes du 16e Congrès International des Linguistes* (CD-ROM), Elsevier.

Tabensky, A. (2001). Gesture and speech rephrasing in conversation. *Gesture, 1*(2), 213–235.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description, Vol. 3: Grammatical categories and the lexicon* (pp. 57–149). Cambridge: Cambridge University Press.

Tao, H. (1999). Body movement and participant alignment in Mandarin conversational interactions. *Papers from the 35th regional meeting of the Chicago Linguistic Society, Vol. II: The panels* (pp. 125–139). Chicago: Chicago Linguistics Society.

Taub, S. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Waugh, L. (1982). Marked and unmarked – a choice between unequals in semiotic structure. *Semiotica, 39, 3/4,* 299–318.

Webb, R. (1996). Linguistic features of metaphoric gestures. Doctoral Dissertation, University of Rochester.

Wilcox, P. P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics, 15*(2), 197–222.

Williams, R. (2004). *Making meaning from a clock: Material artifacts and conceptual blending in time-telling instruction*. Doctoral Dissertation, University of California at San Diego.

Experimental methods for studying language and space*

Laura A. Carlson and Patrick L. Hill

1. Introduction

Research in language and space focuses on the mapping between linguistic descriptions and spatial arrangements of objects, examining which linguistic properties are associated with which spatial properties. For example, consider the spatial description from Talmy (1983) in (1)

(1) "The bicycle is near the house"

The two objects in the sentence serve different roles. The first object (i.e., bicycle) is the located object (also known as target or figure); the goal of the utterance is to assist a listener in finding this object. This is accomplished by spatially relating it to the second object (i.e., the house) that serves as a reference object (also known as a landmark or relatum). Talmy (1983) has argued that properties of the objects in the spatial world are associated with these two roles, with reference objects typically more stable and larger than located objects (see also de Vega et al. 2002; Landau & Jackendoff 1993; Taylor et al. 2000, 2001). This is consistent with the idea that the location of the reference object is presumed known or easily found. This association is best illustrated by contrasting (1) with (2) (also from Talmy 1983). The second utterance is considered less acceptable because the association between the linguistic and spatial properties is violated, with the reference object smaller and more moveable than the located object.

(2) ?"The house is near the bicycle."

Research in language and space can be categorized along a continuum anchored on one end by a focus on identifying the linguistic properties (e.g., Langacker 1993, 2002; Jack-

* We thank the Institute for Scholarship in the Liberal Arts, College of Arts and Letters, University of Notre Dame for supporting Patrick Hill during the writing of this chapter. Address correspondence to Dr. Laura Carlson, 118-D Haggar Hall, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556; email: lcarlson@nd.edu

endoff 1983, 1996; Landau & Jackendoff 1993; Levelt 1996; Talmy 1983) and at the other end on identifying the properties of the spatial representation (e.g., Carlson & van Deman 2004; Eilan et al. 1993; Farah et al. 1990); in the middle is research that focuses on the interface of linguistic and spatial properties (e.g., Carlson-Radvansky & Irwin 1993; Carlson-Radvansky & Jiang 1998; Franklin, Henkel, & Zangas 1995; Garnham 1989; Hayward & Tarr 1995; Levelt 1984; Logan 1995; Mainwaring et al. 2003; Schober 1993; for a review, see Coventry & Garrod 2004). Of interest to the current chapter is research at this interface. Logan and Sadler (1996) describe a computational model of the processes and representations that accomplish the mapping between language and space. For example, consider the scene showing a tube of toothpaste and toothbrush in Figure 1, accompanied by sentence (3):

(3) "The tube of toothpaste is above the toothbrush."

In order to map the linguistic utterance onto the display, the following representations must be created: a representation of the sentence that contains a proposition that assigns roles to the located and reference objects, and that specifies their relation, such as *above*(*toothpaste, toothbrush*); a representation of the picture that contains tokens for the objects; and a conceptual representation (much like a situation model (Tversky 1991; Zwaan & Radvansky 1998)) that links the propositional representation with the perceptual tokens, including associating the object names with their tokens, and the spatial term from the sentence with the spatial relation portrayed in the picture. This latter process involves assigning directions to space, and is accomplished by imposing a reference system on the scene. For projective spatial terms such as "above" or "front", a reference system consists of a set of orthogonal axes that are defined by a number of parameters including the *origin* (where the axes are imposed on the reference object), *orientation and direction* (labeling the axes with respect to the vertical above/below and horizontal left/right and front/back dimensions), *distance* (how space between the objects is defined) and *spatial template* (defining space extending from the axes into good, acceptable and bad regions around the reference object).

Research on the language/space interface has used a number of different methods to examine these processes and representations (for review see Carlson 2003). The purpose of the current chapter is to review a subset of these methods, namely acceptability judgments, speeded verification, placement tasks, tasks that parse space, and production tasks. The following organization is used. For each method, a description is provided that is tied to examples of research that use the method, along with a discussion of illustrative findings. This section also specifies the role of the participant in the task (i.e., as listener or speaker) and the scope of the critical measures (i.e., whether an endpoint measure tied to the final response of the participant, such as acceptability ratings, or an online measure that examines processes unfolding over time, such as event related potentials or eye tracking). Next, there is a discussion of critical design issues for implementing the method. This is followed by a presentation of its strengths, focusing on the kinds of inferences that are licensed, and its weaknesses, focusing on the potential problems that exist and discussing how to confront these limitations. Following the presentation of each method, a final section argues that rather than favoring any given method, what is most impor-

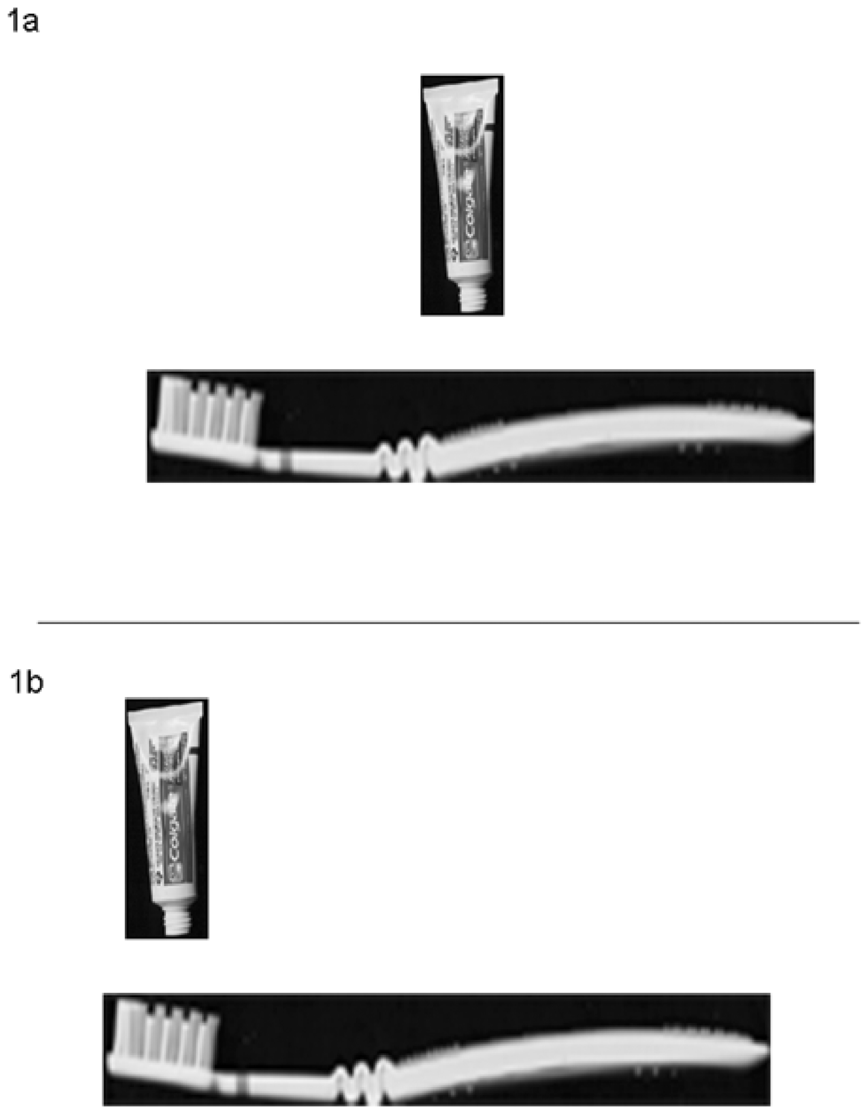


Figure 1. Sample object (toothbrush and toothpaste tube) from Carlson-Radvansky, Covey, and Lattanzi (1999). Panel A. Geometric placement of the toothpaste above the center of mass of the toothbrush; Panel B. Functional placement of the toothpaste above the toothbrush (functional part = bristles).

tant is matching the method to the underlying research question. Moreover, given the different strengths and weaknesses of each method, obtaining a similar pattern of findings across various methods provides converging evidence and gives one more confidence in the obtained pattern.

2. Acceptability ratings

2.1 Description of the method and illustrative studies

In an acceptability rating task, participants are shown objects (real or pictures) in a particular configuration and are provided with an accompanying sentence. The task is to rate the acceptability of the sentence as a description of the picture using a Likert-type scale. Such ratings are an endpoint measure tied to a final response. In this task, the participant is in the role of a listener who makes an assessment regarding the applicability of a given linguistic utterance.

Acceptability rating tasks have been used to assess how different spatial terms are mapped onto space around a reference object (Logan & Sadler 1996; Hayward & Tarr 1995; Carlson-Radvansky & Logan 1997). For example, Logan and Sadler (1996) asked participants to rate how well a variety of spatial terms described the relation between two letters, an X and an O. The O was always placed in the center of a 7 × 7 grid with the X placed across trials in each of the other 48 positions. The set-up of the display is shown in Figure 2; note that participants were not shown the grid. Each sentence took the form "The X is ____ the O", with each of the following ten relations assessed: above, below, over,

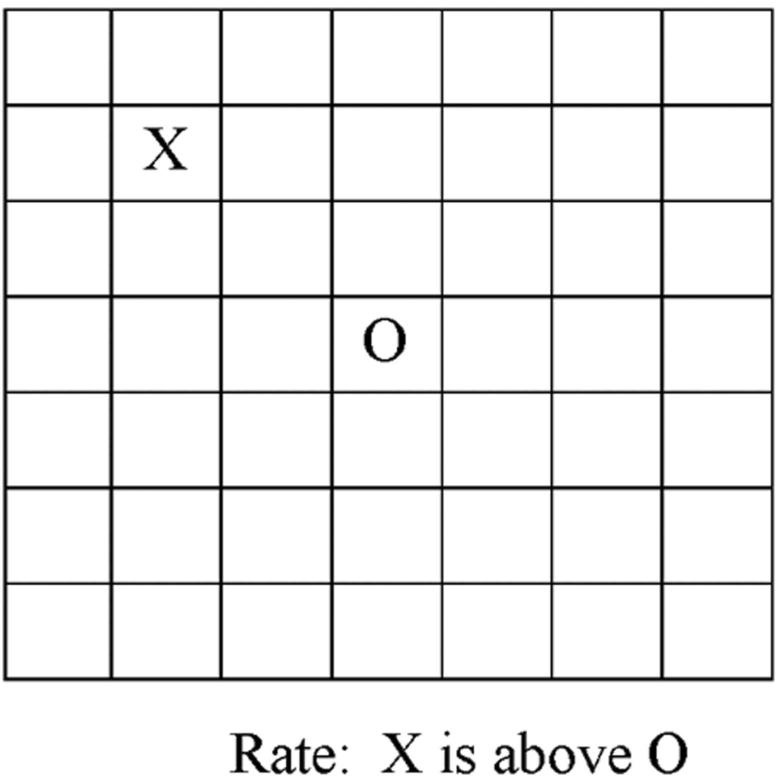


Figure 2. 7 × 7 grid illustrating possible located object placements in the Logan and Sadler (1996) acceptability rating task

under, left of, right of, next to, away from, near to, and far from. Each participant provided a rating of acceptability using a scale from 1 (bad) to 9 (good).

Figure 3 represents the data for the spatial term "above" that result from this type of task; this type of plot has been referred to as a spatial template. The X and Y dimensions define the cells within the grid; the Z dimension corresponds to the acceptability ratings; the reference object (O) was placed in cell (4,4). As seen in the figure, ratings were highest when the X was placed upon a vertical axis running through the reference object, O. This peak has been designated as the good region. Ratings sloped downward as the X moved to the right or the left, but remained moderately acceptable (> 5.5) for rows 1–3 above the reference object. These regions flanked the good region and were designated the acceptable regions. Finally, acceptability dropped considerably when the X was placed in rows 5–7, below the reference object. This region was designated the bad region. Interestingly, Logan and Sadler found that the shape of the spatial template was similar within a given class of terms. For example, spatial templates for projective relations, such as *above*, *below*, *left* and *right*, only differed in their orientation. In addition, proximal terms, such as *next t*o and *near*, were similar, with a peak in the center and a symmetrical drop-off in all directions as X moved away from O toward any of the grid boundaries.

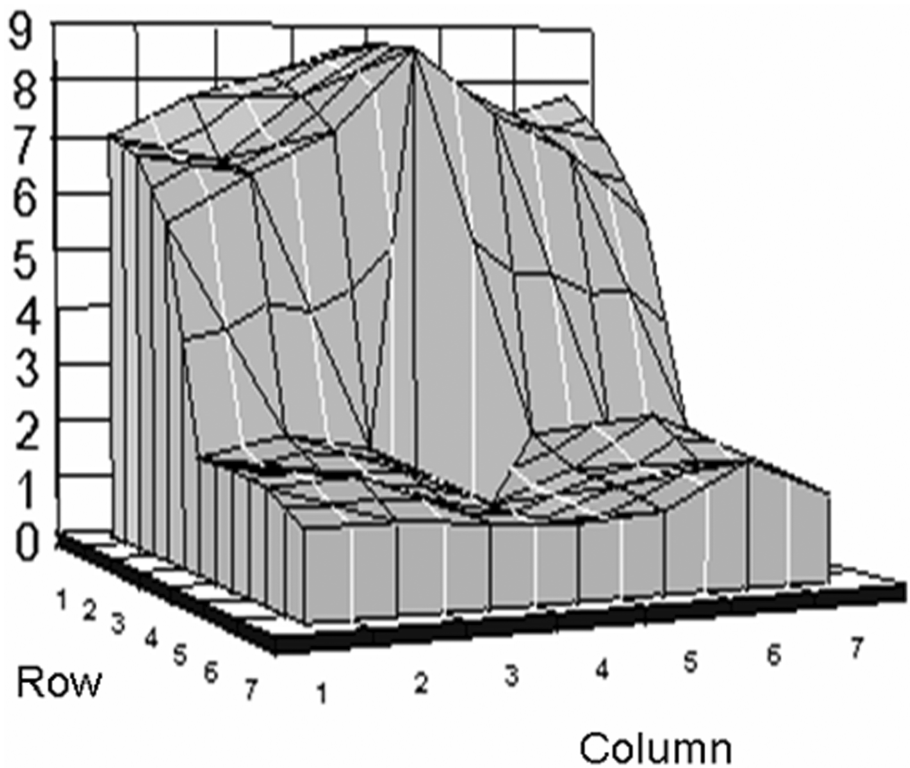


Figure 3. Sample spatial template for "above" based on acceptability rating data from Logan and Sadler (1996).

Acceptability ratings have also been used to determine preference for using different types of reference frames to define spatial terms (Carlson-Radvansky & Irwin 1993; Carlson-Radvansky & Radvansky 1996; Carlson-Radvansky & Tang 2000; Coventry, Prat-Sala, & Richards 2001). There are three general types of reference frames, defined in Part by the source of information used to orient the axes and assign directions to the endpoints (Miller & Johnson-Laird 1976; Levinson 1996). Relative reference frames are based on the participant's viewpoint on the scene; absolute reference frames are based on environmental information such as gravity; intrinsic reference frames are based on the predefined sides of the objects within the scene. For some terms such as "above", these frames are most often in agreement, since participants and objects are often upright and aligned with gravity. Because all of the frames assign the same direction, use of *above* is ambiguous as to which reference frame is being used. For other terms such as *left* and *right*, a dissociation is more typical. For example, for the terms *left* and *right*, speakers and listeners are typically facing each other, such that their sides indicate competing directions for these terms (i.e., the speaker's left is the listener's right).

Carlson-Radvansky and Irwin (1993) had participants rate the acceptability of a given sentence as describing a scene in which the reference frames were dissociated by rotating the reference object. As a result, the intrinsic frame indicated a different direction for *above* than the coincident relative/absolute frames. For example, as seen in Figure 4, the located object (the fly) could either be placed above the reference object (the donkey) with respect to the coincident absolute/relative reference frames (placement 1) or with respect to the intrinsic frame (placement 2). The critical finding was a strong preference for using the absolute/relative reference frame, with a smaller but nonetheless significant use of the intrinsic frame to define the term *above.* In another experiment, participants performed the rating task while reclining on their sides; this manipulation served to further dissociate the relative reference frame from the absolute reference frame. Participants preferred to define *above* with respect to the absolute reference frame, followed by use of the intrinsic frame. There was no evidence of a preference to define this term with respect to the relative reference frame. Finally, acceptability ratings have also shown that the intrinsic reference frame is preferred to the absolute/relative reference frames when describing the spatial relation among objects that are functionally related (Carlson-Radvansky & Radvansky 1996; Carlson-Radvansky & Tang 2000).

2.2 Critical design issues

There are several critical design issues to consider when implementing an acceptability rating task. First, one needs to choose a response scale. This involves determining the number of response categories (i.e., use a scale that runs from 1 to 5 or 0 to 9) (Matell & Jacoby 1971), and whether there should be a midpoint value (i.e., use a scale from 1 to 4 or from 1 to 5) (Guy & Norvell 1977), determining the number of response categories that should be verbally labeled, and determining the labels to use (i.e., not at all acceptable, moderately acceptable, perfectly acceptable). Relatedly, one must also be careful in crafting the instructions to participants to make it clear that the whole range of the scale can be used; use of only the endpoints and the middle effectively reduces the scale to three

Figure 4. Sample stimulus scene with a donkey showing dissociation among reference frames. Reprinted from Carlson-Radvansky and Irwin (1993).

response categories. Moreover, there are cultural issues that may influence how the scale is used, with different groups responding generally more positively or negatively, and using different ranges of values within the scale (Lee, Jones, Minyama, & Zhang 2002).

Second, one needs to be careful in analyzing the ratings, and making inferences about the pattern of responses across conditions. Typically, for these types of experiments, the scales are ordinal rather than interval, because there is no guarantee that the intervals are equally spaced. For example, on a scale using 0 as "not at all acceptable", 5 as "moderately acceptable" and 9 as "perfectly acceptable", the difference between values of 2 and 3 may or may not be the same as the difference between values of 6 and 7. As such, as with any ordinal scale, comparisons across conditions can focus on whether ratings are significantly higher or lower, but can make no statement about the size of the differences that are observed.

Third, one needs to determine which regions of space to probe. For example, one could probe locations that fall only on the axis of the reference frame (within the good region), locations that fall both on and off axes (within the good and acceptable regions), or locations all around the reference object (within the good, acceptable and bad regions).

Moreover, within each region, one could probe finely or coarsely – for example, one could divide the space into a 3 × 3 grid, a 5 × 5 grid, a 7 × 7 grid, a 9 × 9 grid, etc. There is a trade-off with granularity. More fine-grained probing may lead one to believe that a more accurate assessment of the mapping of the spatial term onto space is being obtained; however, more fine-grained probing also requires many more trials (i.e., 8, 24, 48, or 80 for each grid size above). Moreover, if the granularity is finer than the parsing of space that the participant would normally make when using these terms, this may invite an implicit contrast across trials that could lead to artificial distinctions.

2.3 Strengths and weaknesses

2.3.1 *Strengths*

One significant strength of an acceptability rating task is that it offers a fine-grained measure which provides more information than a simple binary "yes" or "no" decision (see Verification tasks) which does not provide any allowance for more marginal uses of the spatial term. This is a benefit because many theorists have argued that spatial terms correspond to regions of space around the reference object instead of corresponding to a solitary point (Hayward & Tarr 1995; Langacker 1993, 2002; Miller & Johnson-Laird 1976; Morrow & Clark 1988); thus, assessing how the term is mapped across points within these regions is appropriate for questions that compare these regions across different spatial terms, across different types of objects, or across levels of whatever manipulation is of interest to the research.

Another benefit is that rating tasks resulting in spatial templates provide rich data sets that can be used in more formal computational approaches to defining spatial terms. For example, Regier and Carlson (2001) developed the attention vector sum (AVS) model to account for the similarity in spatial template shape across the projective relations (Logan & Sadler 1996). This model involves an attentional beam that is centered on the reference object and radiates outward to encompass the located object. In addition, direction is represented as a sum of vectors anchored within the reference object, and pointing toward the located object, with the relative importance of each vector weighted by the amount of attention allocated to that point. This model provided a better quantitative fit to the human rating data than competitor models, and qualitatively exhibited the effects in its output.

2.3.2 *Weaknesses*

Several potential weaknesses are associated with acceptability ratings. First, the rating is in some ways an artificial measure that does not transparently map onto actual use. For example, for a term receiving a rating of 6 on a scale from 1 (not acceptable) to 9 (perfectly acceptable) as a description of an arrangement of objects, it is not clear whether this rating indicates that a participant would actually use the given term in natural speech to describe the location of an object, or whether an alternative term would be preferred. Second, participants may have preferences for using only the endpoints or middle values of the scale, with a bias against using a full set of response values, or different participants may use the scale differently. Third, the space that can be probed is constrained to the edges of the display, typically within a computer monitor. This limits the ability to treat the spatial template as a direct representation of the mapping of the term onto space (see Parsing Space Tasks). Fourth, ratings tasks may set up implicit contrasts across the trials, such that responses on one trial may be biased by the rating assigned to an object in a neighboring cell on a preceding trial. Finally, thus far, use of this task has been largely limited to assessment of the relation between the target and the reference object in scenes containing only these objects (e.g., Hayward & Tarr 1995; Logan & Sadler 1996). However, rarely are objects in the real world isolated from other objects. Moreover, the presence of additional objects may well impact judgments of acceptability (Herskovits 1986). Recently, Carlson and Logan (2001) used a combination of rating and verification tasks to address this issue.

3. Speeded verification

3.1 Description of the method and illustrative studies

In a speeded verification task, participants are shown pictures of objects in a given configuration and are provided with a sentence. The task is to determine whether the sentence is an acceptable description of the picture. The response time and accuracy for making this judgment constitute the critical dependent variables. In this task, the participant is in the role of a listener, assessing a linguistic description that is provided to them. In addition, this paradigm uses an online measure that allows one to make inferences about the relative difficulty across conditions, as reflected in the pattern of response times.

Speeded sentence/picture verification tasks have been used in a variety of studies to address a variety of questions (Carlson-Radvansky & Irwin 1994; Carlson, West, Taylor, & Herndon 2002; Carpenter & Just 1975; Clark & Chase 1972; Seymour 1973). For example, Carlson-Radvansky and Irwin (1994) used this paradigm to demonstrate that the selection of a reference frame for defining spatial terms involved an initial period in which multiple frames were active simultaneously and competed for selection. One way in which this was demonstrated was by comparing "no" response times in two critical conditions: responding no to a placement that was above a reference object with respect to the less preferred intrinsic reference frame (i.e., Placement 2 in Figure 4) versus responding no to placement that was not above the reference object with respect to any reference frame. Response times were significantly slower for rejecting the trials in which the located object was placed intrinsically above the reference object, indicating an initial activation of the intrinsic reference frame. Such activation could not have been observed using other methods in which only the final response was recorded (for example, acceptability ratings).

Speeded sentence/picture verification tasks have also been used to demonstrate that information about the function of the objects being spatially related and how they interact affects the use of spatial terms (Carlson, Regier, Lopez, & Corrigan 2004). The speeded sentence picture verification task has also been combined with event related potentials (Carlson, West, Taylor, & Herndon 2002) in order to obtain independent evidence of the underlying processes governing the use of spatial terms (for other uses of ERPs within language, see the chapter by Coulson this volume).

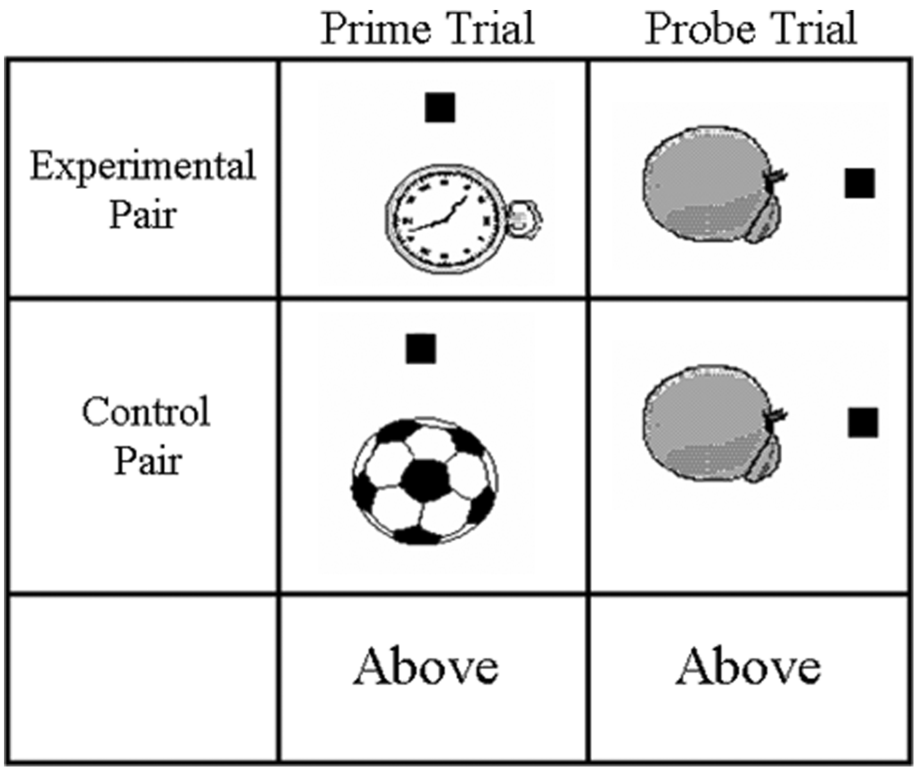


Figure 5. Illustration of the prime and probe trials within the negative priming verification paradigm. Reprinted from Carlson-Radvansky and Jiang (1998)

Importantly, one of the powerful aspects of the sentence/picture verification paradigm is flexibility with which a trial stream can be constructed. It is relatively easy to set up a systematic relationship among pairs of trials, allowing one to assess how processing on one trial affects processing on a subsequent trial (Carlson-Radvansky & Jiang 1998; Carlson & Van Deman 2004). For example, Carlson-Radvansky and Jiang (1998) used a negative priming paradigm to assess whether the selection of a reference frame (absolute/relative) involved inhibition of an active but non-selected frame (intrinsic). Participants are instructed to respond "yes" to both absolute/relative and intrinsic placements of a located object; thus, for example, for all trials in Figure 5, the response is "yes." Two types of prime/probe pairs of trials were created, an experimental pair and a control pair. The probe trials were identical across both of these pairs of trials. This is illustrated in Figure 5, showing a rotated peach as a reference object, and a small square placed intrinsically above the peach. Because the pictures, the sentence, and the response on these trials are identical, any difference in response times between experimental and control probe trials must be due to processing differences on the prime trial. Consider next the experimental prime trial in Figure 5 with a rotated pocket watch and a small square placed above it with respect to the coincident absolute/relative frames. Note that because a pocket watch has an intrinsic top side, it is possible to define above with respect to the intrinsic frame.

However, on this trial, participants are defining above with respect to the absolute/relative reference frame. The critical question is whether the intrinsic frame that is presumed active on this trial (Carlson-Radvansky & Irwin 1994) is inhibited during the selection of the absolute/relative reference frame. This is assessed by contrasting this trial with the control prime trial that shows a soccer ball with a small square placed above the ball according to the coincident absolute/relative frames. Because balls do not have intrinsic top sides, there is no possible use of the intrinsic reference frame on this trial, and thus no potential for inhibiting it. Therefore, the logic was that if the intrinsic frame is inhibited on experimental prime trials but not on control prime trials, use of the intrinsic frame should be more difficult on experimental probe trials than on control probe trials; this should be revealed by slower response times to the experimental probe trials than the control probe trials. This pattern was observed, and indicated that inhibition played a role in the selection of a reference frame.

Another example of using a systematic relationship between prime/probe trials to assess underlying processing comes from Carlson and van Deman (2004) who examined whether the distance between two objects being spatially related was encoded on a given trial. In that study, displays consisted of pairs of letters, with accompanying sentences such as "L is above X." Across a pair of prime/probe trials, the identity of the letters and their placement on the screen varied. The critical manipulation was whether the distance between the two letters matched or mismatched across primes and probes. Sample displays are shown in Figure 6. The logic was that if participants encoded the distance between the letters on the prime trial, then responding on the probe trial when the letters were at the same distance should be facilitated, compared to when the letters were at a different distance. Such facilitation was observed when participants verified sentences that spatially related the two letters. However, using the same displays and paradigm, when participants verified sentences using an *and* relation (L and X), no such facilitation was observed. This was taken as evidence that distance is relevant and encoded during the processing of spatial terms, even for terms such as *above* that do not explicitly convey distance as part of their definition (as compared to *near* for instance).

3.2 Critical design issues

There are several critical design issues for speeded sentence/picture verification. First, one must determine whether to present the sentences and pictures simultaneously or sequentially, and if sequentially, whether to present the sentence first or the picture first (for discussion see Carpenter & Just 1975, 1976; Catlin & Jones 1976; Clark & Chase 1972; Liu 1980; Shoben 1978; Singer 1977). Moreover, different strategies may be adopted by different participants, with some focusing on a more linguistic strategy and others on a more image-based strategy (Glushko & Cooper 1978; MacLeod, Hunt, & Matthews 1978; Reichle, Carpenter, & Just 2000). The sequence of stimuli, the time course of presentation (i.e., how much time to allow for processing the sentence and the picture (Tversky 1975), and the instructions to participants should take into account the potential for these individual differences. Second, the composition of trials should be constructed with care. It may be best to equate the overall number of trials requiring "yes" and "no" responses;

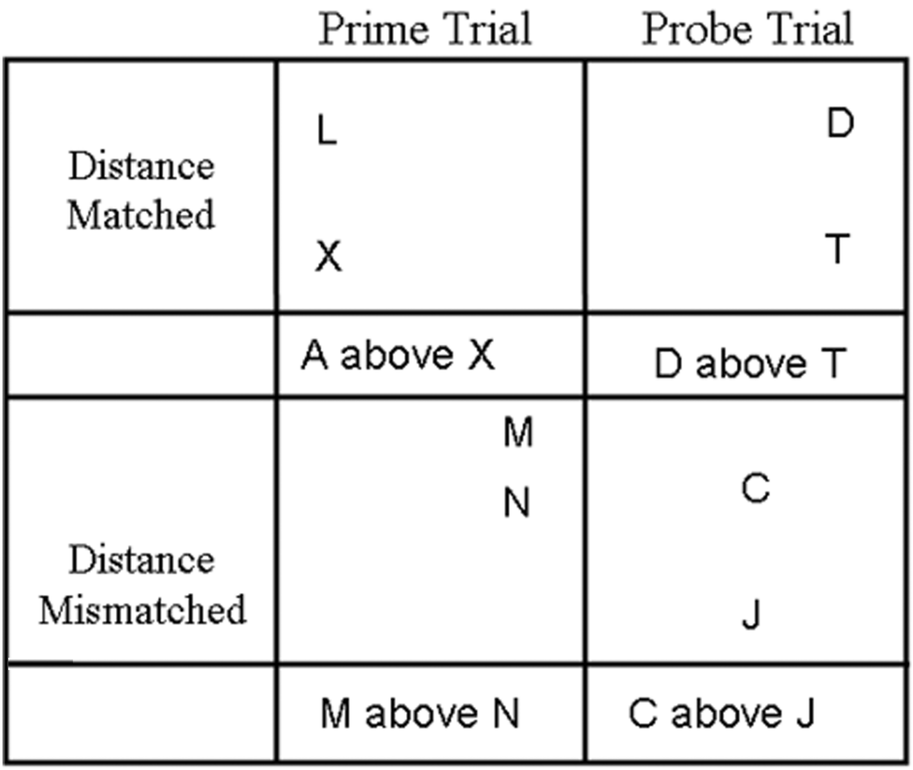


Figure 6. Sample prime and probe trials used by Carlson and van Deman (2004), showing distance matched and distance mismatched pairs.

this helps eliminate any decision bias. Note that this tends to greatly inflate the number of possible trials. Moreover, it is difficult to directly compare "yes" and "no" responses; "no" responses typically take longer than "yes" responses, and it is not always clear why this is the case, or whether the processing underlying these judgments is the same (for discussion, see Clark & Chase 1972; Carpenter & Just 1975, 1976). As such, if the conditions of interest all require "yes" responses, the "no" responses are treated largely as fillers. Note, however, some important questions can be answered by comparing across conditions within "no" responses (see for example, Carlson-Radvansky & Irwin's (1994) comparison of "no" trials that signaled competition between active reference frames). If there is theoretical motivation for these comparisons, the "no" trials can be very valuable. Third, different types of "no" trials should be used in order to prevent the use of strategies. There are at least three possible types: 1) trials in which the names of the objects in the sentence do not match the objects in the picture, with some trials with an incorrect correspondence for the located object, some with an incorrect correspondence for the reference object, and some with an incorrect correspondence for both objects; 2) trials in which the spatial term in the sentence does not correspond to the spatial relation between the objects in the picture; and 3) trials in which there is an incorrect correspondence for both the names of the objects and the spatial term. Fourth, it is usually advisable to have an initial set of object identification trials in which the objects to be used in the pictures are presented to the participants, along with the label that will be used to refer to them. This will minimize potential effects due to difficulties in recognizing the objects that could inflate verification times, particularly if there are manipulations that render the objects more difficult to identify (such as presenting them in a noncanonical orientation, or using a set of stimuli for which no naming norms have been established). Fifth, one needs to determine whether to present the trials in a self-paced fashion under the control of the participant or in a continuous stream under the control of the experimenter. The advantage of self-paced trials is that participants can take breaks whenever they want, and this is advantageous given the large number of trials that are typically involved in these types of experiments (i.e., 300–600). However, some paradigms require precise inter-trial timing, such as the prime/probe paradigms. For these types of experiments, the trials can be split into blocks, with breaks built in between blocks. In such cases, it is beneficial to have factors counterbalanced within a block so that one can assess whether effects of interest vary across blocks as a function of learning, participant fatigue, and so on. Finally, comparing response times across conditions invites the use of the subtractive method, where differences in response time are attributable to differences in the set of processes that are active across the conditions. For example, if condition A uses processes 1, 2 and 3 and condition B uses processes 1 and 3, then the difference in processing time between A and B can be attributed to process 2. Several assumptions must hold for this logic to work. For example, processes 1 and 3 must operate exactly the same in conditions A and B. Also, process 2 must not overlap in time with processes 1 and 3; otherwise, the estimate of time due to the difference between A and B will underestimate this process (for full discussion, see Sternberg 1998).

3.3 Strengths and weaknesses

3.3.1 *Strengths*

There are a number of strengths associated with speeded sentence/picture paradigms. First, they provide the experimenter with two sets of data: the response that was generated (yes or no) and the time required to make that response. Second, response times can be used to index particular cognitive processes that may not be observable by endpoint responses such as ratings or yes/no judgments. For example, Carlson-Radvansky and Irwin's (1994) finding that multiple reference frames compete for selection could not have been observed in a task that only looked at which frame was ultimately selected. Third, the paradigm is flexible and can be adapted for a wide range of research questions. There are multiple ways to implement a manipulation, including changing the stimuli in a picture, changing the sentence, and changing the type of judgment the participant is making. In addition, it can be combined with other indirect measures such as eye tracking (Just & Carpenter 1976) or event related potentials. For example, Carlson, West, Taylor, and Herndon (2002) found distinct neural correlates (P3, frontal slow wave, parietal slow wave) associated with the processes underlying the comprehension of a spatial description (spatially indexing and identifying the objects; assigning directions in space; computing and comparing the spatial relation).

3.3.2 *Weaknesses*

There are several weaknesses to speeded sentence/picture verification tasks. First, as with acceptability rating tasks, it is an artificial measure that requires interpretation on the part of the experimenter as to what factors or change in processing resulted in observed differences across conditions of the experiment. This may prove difficult, as changes in response time could also be due to extraneous factors such as difficulty in identifying the objects, reading the sentence, lapses of attention, and so on. Given a within subject design, these effects should be present in all conditions; however the added variability may make it difficult to detect a systematic effect due to an independent variable of interest. As such, multiple trials should be obtained within each cell, and this inflates the number of trials. Second, due to the potential for variability in this measure, a large number of trials are required. Third, if the set of stimuli is limited, the large number of trials is achieved by repeated presentations; familiarity with the stimuli over the course of the experiment could alter the way in which the participants are performing the task and the underlying processes that are involved. Alternatively, one could use a large stimulus set with few repetitions, but often there is considerable labor involved in creating such stimuli.

4. Placement tasks

4.1 Description of the method and illustrative studies

In a typical placement task, participants are given props (real objects, pictures of objects) and asked to place one object in a specified relation to another object, as in (4):

(4) "Put the coin above the piggy bank."

The participant is in the role of a listener, although they are using the language to accomplish a particular goal: arrange the objects so that the utterance is true. Of interest are the resulting arrangements of objects. Carlson-Radvansky, Covey and Lattanzi (1999) used a placement task to show that the origin of the reference frame was defined by geometric and functional information. Traditionally, spatial terms have been defined geometrically, with the axes of a reference frame imposed on the presumed center of the reference object, and locations along the axes corresponding to the best use of the spatial term. This is represented by the placement of the toothpaste tube above the center of the toothbrush in Figure 1a. Within this account, the object is schematized as an abstract shape with the axes preserved and properties of the reference object such as its function or use are considered irrelevant (Herskovits 1986; Landau & Jackendoff 1993; Talmy 1983). However, given that artifacts are created to fulfill a function, and that objects are often placed in a spatial relationship to fulfill that function, Carlson-Radvansky et al. (1999) reasoned that it was possible that the origin could be defined by a relevant functional part of the object, rather than its geometric center. This possibility is illustrated in Figure 1b, in which the toothpaste is above the bristles of the toothbrush.

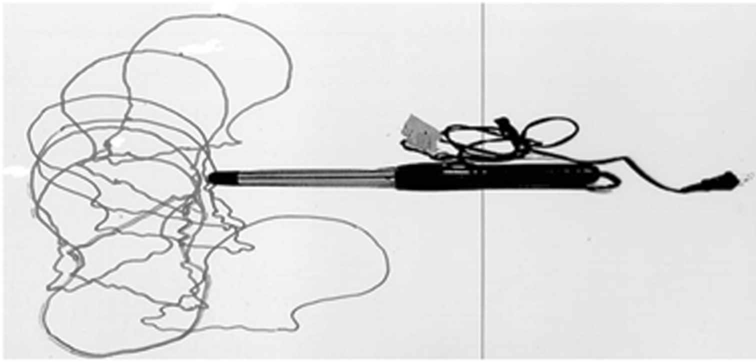
To assess the role of geometric and functional information, Carlson-Radvansky et al. (1999) took pictures of references objects from a perspective in which an important functional part was offset from the center of the object, as in Figure 1 with the bristles of the toothbrush on the left. Participants were given pictures of objects that typically interacted with the reference objects, and asked to place them above or below the reference object. For example, for the toothbrush as the reference object, participants would be asked to place the tube of toothpaste above the toothbrush. The placement of the target constituted the main dependent variable. Importantly, the functional parts of the objects (i.e., bristles) were never explicitly mentioned. Nevertheless, the critical result was that placements were biased away from the center of the object and toward the functional part. Recently, Carlson and Kenny (2006) extended these results by showing that the part of the reference object that was deemed functional was determined by the interaction between the located object and the reference object. That is, many objects have many functional parts, and the parts may be differentially salient as a function of how the object is being used. As such, it is possible that the functional bias observed by Carlson-Radvansky et al. (1999) may be defined with respect to different functional parts. To assess this, they selected objects with two functional parts on opposite sides of the object, and selected different located objects, one that interacted with each part. The task was to place the located object in some spatial relation to the reference object; again, the functional parts were not mentioned. Representative placement data for the term "near" for a curling iron as the reference object are shown in Figure 7, with a wig (interacts with the iron end, see Figure 7a) and a power strip (interacts with the plug, see Figure 7b) as the located objects.

In a different use of the placement task, Chambers et al. (2002) asked participants to manipulate objects within a visual workspace in accordance with various instructions, such as placing one object inside another object. Participants wore a head-mounted eye tracker while performing the task. Eye tracking has recently been developed as an important tool for studying language comprehension (Tanenhaus et al. 1995; see also Eberhard et. al 1995; Spivey et al. 2002), particularly with respect to establishing reference. For example, in response to the instruction "Put the cube inside the can", participants need to link the term "can" with the correct entity in the display. Chambers et al. found that participants were able to restrict the objects that are considered as intended referents for "can" to those that supported containment. This suggests an online integration of information from the visual display and semantic-conceptual constraints imposed by the spatial term. This study nicely illustrates the benefit of combining two different methods (placement task and eye tracking) within a single study (see also Richardson, Dale, & Spivey this volume on the use of eye tracking as a measure).

4.2 Critical design issues

There are several critical design issues for placement tasks. First, quantifying the placement is non-trivial. For example, in Carlson-Radvansky et al. (1999), the placements were coded as deviations from a line running through the center of mass of the reference object, with placements toward the functional part designated as a positive deviation and placements away from the functional part designated as a negative deviation. Thus, the sign as well as the value of the deviation provide important information. Second, determining the best dimension for calculating the deviation may not be trivial. For terms like

7a



7b

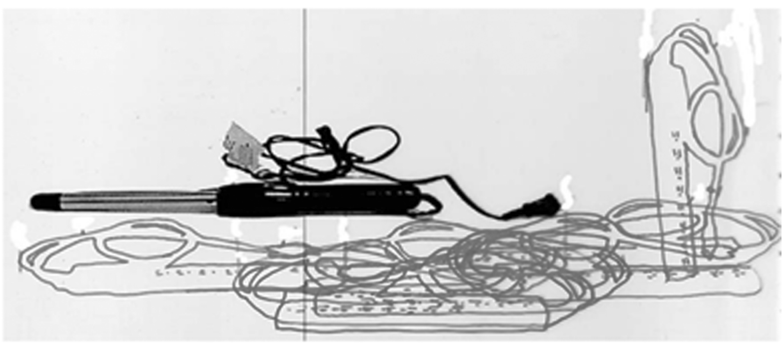


Figure 7. Sample objects from Carlson and Kenny (2006) used in the placement task. Reference object is the curling iron. Panel A shows placements of the wig around the iron (the relevant functional part). Panel B shows placements of the power strip around the plug (the relevant functional part).

"above" or "below" horizontal deviations from a vertical line running through the center of the object would be most interpretable. However, it is also possible that interesting differences exist in the vertical dimension for these placements, and this should not be overlooked. Moreover, for terms like "near" that do not convey a direction, it is likely that both vertical and horizontal deviations would need to be taken into account, both individually and in combination. Third, the measure should be converted into a standardized value that will allow comparisons across objects. For example, for the reference objects used by Carlson-Radvansky et al. (1999), there was a large variation in how far away the functional part was from the center of the object. Thus, coding deviations in terms of millimeters was uninformative across objects as to the degree of bias. For example, a deviation of +3 mm for an object with a distance of 4 mm between the functional part and the center would be considered a relatively large deviation, whereas the same deviation for an object with a distance of 9 mm between the functional part and the center would be considered a relatively small deviation. Accordingly, in Carlson-Radvansky et al., deviations were converted into percentages, taking the distance into account. Within this scheme, the deviation for the first object would be + 75% whereas the deviation for the second object would be +33 %. Fourth, determining dimensions of the area (the page, the display) where participants are making their placements, relative to the size of the object, is important. One needs to make sure that sufficient space is available to allow for variation, should it be present; at the same time, too much space may encourage participant to artificially determine object placements (either close or far from the reference object). This is far from settled. Finally, an important issue is whether the design is within or between subjects. Within subject designs have the advantage of allowing one to interpret the data pattern across conditions for a given subject, thereby taking idiosyncratic preferences or strategies into account. However, within subject designs also may invite participants to make explicit comparisons across conditions, thereby biasing placements. The disadvantage of a between subjects approach is that differences across the conditions could be due to participant strategies as opposed to the manipulations under investigation.

4.3 Strengths and weaknesses

4.3.1 *Strengths*

There are several strengths of the placement task that are not present in the rating or verification tasks. First, placement tasks are more naturalistic, and use language to accomplish a goal. The linguistic utterance is meaningful in this regard, unlike the process of matching the sentence to picture in the rating and verification tasks in which the goal is somewhat nonlinguistic (matching). As such, this task is more consistent with a view of language as an active process involving two or more parties that have to coordinate activities in order to accomplish a goal (Clark 1996). Second, each placement is informative as to how the participant is mapping language onto space, unlike verification filler trials that are typically discarded. Finally, the task is more general in that it taps into additional cognitive systems, such as the action system. The inclusion of the action system is consistent with recent work that takes an embodied cognition approach to language comprehension (e.g., Glenberg & Kaschak 2002; Zwaan 2004).

4.3.2 *Weaknesses*

There are several weaknesses associated with the placement tasks. First, different types of conclusions can be drawn, depending upon whether analyses focus on objects, collapsing across participants, or focus on participants, collapsing across objects. Focus on the objects would make sense in that different objects have different functional parts with differential salience. For example, knowing how *above* is defined with respect to a toothbrush may not be very informative as to how *above* would be defined with respect to a curling iron. Thus, preserving the effects at the object level would facilitate interpretation. However, at the same time, participants have different strategies, and some may be more biased to use geometric information and some may be more biased to use functional information, independently of the objects being tested. Thus preserving the effects at the subject level would facilitate interpretation. Certainly both types of analyses should be done. Second, for a placement task that uses pictures, the researcher must decide whether to preserve the relative size of the objects within a scene or the absolute size of objects across scenes. For example, if objects are scaled to fit in a center three degree square of a page, a pocket watch on one trial and a house on the next trial will have the same absolute size, but their relative sizes will not be accurately displayed. This makes it difficult to infer how participants are conceptualizing the space around the objects. However, without equating the size of the objects across trials, it becomes difficult to make comparisons of placements across objects. Third, it is potentially difficult to interpret differences in the magnitude of any effect that is observed across objects. For example, a larger functional bias for one particular object (such as the piggy bank) than for another object (such as the hammer) could be due to multiple factors, such as the presence of other functional parts on an object that exert a bias in another direction (for example, the handle and the head of the hammer located on opposite ends), or the fact that the strength of the functional relationship was weaker for a given object. Finally, although a placement task uses language to accomplish a particular goal (i.e., arrange objects so that the utterance is true), it is not necessarily the case that the reverse is true: that given a particular placement of the objects, that the same spatial term may not always be used to describe the relation.

5. Parsing space tasks

5.1 Description of the method and illustrative studies

In a parsing space task, participants are instructed to parse the space surrounding an object into regions, with the focus on defining dimensions of these regions including their boundaries and spatial extent. For example, Franklin, Henkel and Zangas (1995) used a parsing space task to determine the boundaries that corresponded to the *front*, *back*, *left* and *right* regions of space around themselves. Specifically, participants stood inside a uniform circular room that was surrounded on the outside by a ruler that specified angles that was invisible to the participants. They faced a given orientation, and were asked to point as far in one direction as possible while still pointing within a given region. The position of the pointer was recorded, and the difference between it and the starting location was an index of half of the front region. For example, a participant could start at an orientation corresponding to 70 degrees. To assess the left/front boundary, a participant would be asked to move the pointer as far left as possible while still pointing to the front. If the final pointer position was at 10 degrees, then the front region from midpoint to the left edge would extend 60 degrees. Using this paradigm they found that rather than the space being equally divided into 90 degree quadrants, the *front* and *back* regions were largest (124 degrees and 110 degrees, respectively), followed by the right and left regions (92 degrees and 91 degrees, respectively).

In a complementary approach, rather than focusing on assessing the boundaries of *front*, Carlson, Murray, Shreiner & Carolan (2004) used a parsing space paradigm to assess the spatial extent of the *front* region. Specifically, they placed a 5 cm (width) × 7 cm (length) × 24 cm (height) object (i.e., a doll cabinet) in the middle of a 102 × 82 cm uniform white board. This is shown in Figure 8a. Eleven lines radiated out from the object, with lines 5–7 corresponding to the good region of a spatial template, lines 2–4 and 8–10 corresponding to the acceptable region, and lines 1 and 11 corresponding to the bad region; these are shown in Figure 8b. Participants were asked to provide two direct measures of distance: one indicating the best location of *front*, and one indicating the farthest location on the line away from the object at which they would still be willing to use *front*. These regions are shown in Figure 8b, with the best *front* shown in dark and the farthest *front* shown in light gray. Using this paradigm, Carlson et al. (2004) showed that the size and shape of these regions are significantly impacted by the identity of the objects, the reference frame defining the term, and the presence of additional objects.

5.2 Critical design issues

The critical design issues for a parsing space method center upon constructing the space that will be measured. In order to measure space around oneself as in Franklin, Henkel and Zangas (1995), a uniform circular room must be constructed, with no visible landmarks from which to judge direction. For measuring the spatial extent of regions denoted by a spatial term such as *front*, sufficient space around the object must be provided to allow participants the freedom to construct as large a region as desired. In addition, other objects and landmarks must be removed from view that could be used to scale the distance. Finally, it is not always clear which set of instructions will best reveal the spatial extent of the region, and how that will map onto actual linguistic use of the term. For instance, Carlson et al. (2004) asked participants to indicate both the "best" front and the farthest "front." The inference is that participants would then find it appropriate to use the term "front" to describe objects placed within these regions in a production task; however, this has yet to be established.

5.3 Strengths and weaknesses

5.3.1 *Strengths*

A strength of the parsing space method is that offers a direct measure of the mapping of language onto space. Participants are asked to indicate the boundaries of a region associated with a spatial term or the spatial extent of the region, and their responses serve to define the region in actual space. In addition, there is also little ambiguity as to the goal for the participant; they are simply asked to indicate how space around themselves or around an object is parsed.

8a

8b

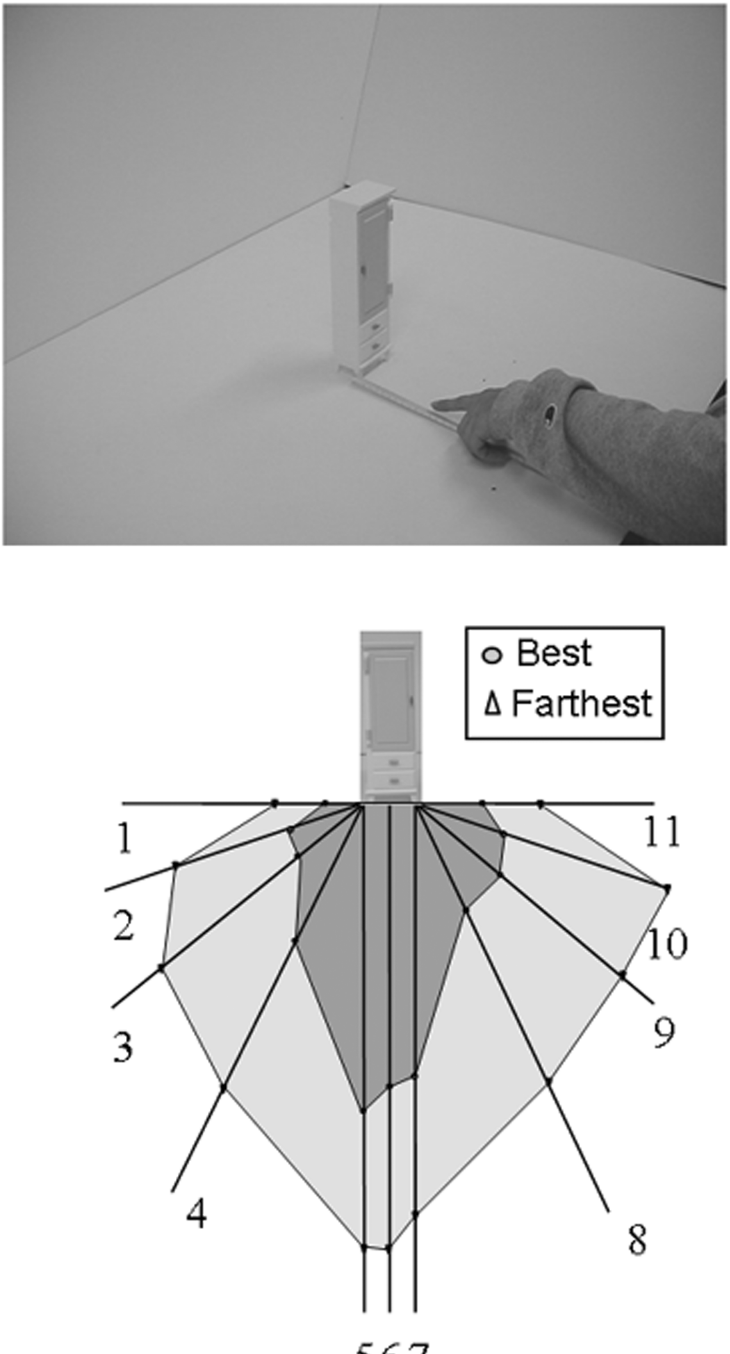


Figure 8. Central object (doll cabinet) used in the parsing space method from Carlson, Murray, Shreiner and Carolan (2004). Panel A shows experimental set-up with display board. Panel B shows regions corresponding to the best "front" (in dark gray) and the farthest "front (in light gray).

5.3.2 *Weaknesses*

A weakness of the parsing space method is that the space that is being considered is functionally limited to the space that the experimenter constructs around the object, including the size of the room around the participant or the size of the board on which the object is placed. In addition, it is often unclear how participants are parsing the space that is available to them. For example, contrast the placement of a small object (i.e., lotion bottle) with a size-matched model object (i.e., dollhouse dresser). It is not clear whether participants will encode the space around the objects in similar size units, or scale them on the basis of the objects' status as a model or actual size object. In addition, as in the placement tasks, it is difficult to determine whether a given factor should be manipulated between participants or within participants. The within participant design invites a contrast across conditions that may elicit an unwanted strategy; however, a between participant design does not enable one to control for individual preferences in the size of these regions. Finally, these types of parsing space tasks are relatively new, and many of the methodological issues are still being worked out.

6. Production

6.1 Description of the method and illustrative studies

In a production task, participants are shown a configuration of objects and asked to offer a description. The task can be fixed, in that their productions are drawn from a limited set, or open-ended, with no restrictions on their productions. For an example using a fixed set of utterances, Carlson-Radvansky and Radvansky (1996) provided participants with scenes containing two objects, and sentence frames of the form Object1 is ________ Object 2. Participants were asked to select the spatial term that best described the relationship between the objects. For some pictures, the objects were arranged to permit a typical functional interaction, such as a mail carrier facing a mailbox. For other pictures, the same two objects were arranged in a way that violated the typical interaction, such as a mail carrier facing away from the mailbox. For each scene, at least two terms were possible, one defined with respect to the intrinsic reference frame defined by the reference object, and one defined by coincident relative/absolute reference frames defined by the participant and the picture environment. For example, the mail carrier in both pictures could be described as being either in front of the mailbox (use of the intrinsic frame) or to the left of the mailbox (relative/absolute frame). Consistent with theoretical suggestions that the intrinsic frame emphasized the functional relationship among the objects (Miller & Johnson-Laird 1976; Ullmer-Ehrich 1982), participants preferred to use the intrinsic frame when the objects were arranged to permit the functional interaction but the relative/absolute frame when objects were arranged in a way that prohibited the interaction. Recently, Richards and Coventry (2004) found the same pattern of findings using an open-ended version of this task.

In contrast to fixed sets of utterances, open-ended tasks allow the speaker more freedom in formulating their utterances. For example, Plumert, Carswell, De Vet, & Ihrig

(1995) asked people to describe learned locations of hidden objects in a hierarchical space (a dollhouse) in which locations were designated at multiple levels (i.e., spatially relating the target to small and large reference objects in a particular room on a particular floor in the house). Plumert et al. were particularly interested in the structure of these utterances. They observed two distinct patterns that corresponded to the instructions given to participants. Specifically, when instructed to provide directions to find the object, utterances typically followed a descending order in which speakers started with the most general spatial unit, and worked toward the most specific one (i.e., "the salt shaker is in the kitchenette on top of the fridge in the popcorn bowl"). In contrast, when instructed to describe the location of the object, utterances typically followed an ascending order in which speakers started at the most specific spatial unit and worked toward the most general (i.e., "the salt shaker is in the popcorn bowl on top of the fridge in the kitchenette"). Other important open-ended production tasks include Schober's (1993, 1995) studies in which directors described to matchers the arrangement of objects on a circular "pizza"; Linde and Labov's (1975) study of apartment descriptions; Taylor and Tversky's (1996) examination of perspective in route and survey descriptions; Mainwaring et al.'s (2003) spy scenario for eliciting spatial descriptions and the HCRC map task and corpus (Anderson et al. 1991).

6.2 Critical design issues

Critical design issues in production tasks differ depending upon whether the responses are fixed or open-ended. If using a fixed set of alternatives, one important issue is determining which response options to include; the trade-off is between offering enough flexibility for participants to have some choice, while constraining the options so that sufficient data in the conditions of interest can be collected. One consequence of a fixed-set approach is that one is not certain whether the option that is selected by the participant corresponds to the one that the participant would use in an open-ended task. With respect to using an open-ended task, the critical issue would be designing the stimuli and the task in such a way that the data will be informative as to the particular manipulations under examination. For example, if one wants to determine whether a speaker will adopt the listener's perspective, one needs to make sure that there are conditions in which the speaker's and listener's perspectives are offset from each other in order to determine whether perspective has an impact on what is said. However, this approach sometimes leads to an ad hoc cataloging of differences across conditions with little understanding as to the causes of these differences. As such, this method is beneficial for describing what participants do say, but it may be difficult to move from that level of description to be able to argue about why they say what they say.

6.3 Strengths and weaknesses

6.3.1 *Strengths*

As outlined by Anderson et al (1991), there are many strengths of an open-ended production task. First, there is a clearly defined goal that the speaker and listener must accomplish, and one can study how subjects interact to accomplish this goal, as well as examine the strategies that are used (e.g., Brennan & Clark 1996). Second, successful communication can be measured as a function of whether the goal was accomplished, allowing one to compare properties of the utterances that seem to lead to different rates of success. Third, the domain of discourse is defined independently of what is said. For example, within the map task, the set of possible landmarks is defined on the map. Thus the research can identify not only which landmark the participant chose to mention in his/her utterance, but also which ones were not used. Finally, the language that is being examined is almost solely in the control of the subject and not the experimenter, and thus has naturalistic appeal.

6.3.2 *Weaknesses*

One prominent weakness of an open-ended production task is a consequence of the unstructured nature of the task. That is, by relinquishing control over the language used, researchers can no longer be certain that particular types of utterances will be produced. This is a problem only if particular types of utterances are central to hypotheses being tested. In a close-ended task, experimenters are in more control of the language that will be used. However, by imposing restrictions on what is permitted, researchers cannot be certain that their findings correspond to natural speech. This is a difficult trade-off to negotiate. Second, production tasks involve detailed transcription of the speech produced by the participants, and this can be a time-consuming process. One possibility is to use existing corpora that have already been transcribed and coded with respect to various linguistic properties. For one example, the map task corpus from Anderson et al. (1991) can be found at www.hcrc.ed.ac.uk/dialogue/maptask.html.

7. Combining experimental methods: Converging measures

This chapter has covered a diverse set of methods that have been used to examine the interface between language and space. These include acceptability ratings, verification, placement tasks, parsing space, and production. In addition, studies combining these methods with computational modeling, event related potentials, and eye tracking were briefly described as a way to illustrate the benefits that pertain to combining multiple measures within a single study. Significant benefits are also obtained by combining methods across studies in a programmatic line of research. For example, in conjunction with colleagues, we have used a combination of acceptability ratings, verification tasks, and event related potentials to study the processes underlying the selection of a reference frame. A combination of methodologies has also been used to examine how functional information about an object impacts descriptions of its location, including rating, placement, and parsing space tasks. Given that each method has its own strengths and weaknesses, the best approach is to use a combination of methods, such that the strength of one offsets the weakness of the other.

Each of these methods is a valuable tool for addressing some aspect of the language and space interface. As such, we resist offering a rank ordering of the methods in terms of importance, general utility or any other dimension. Instead, we believe that it is more beneficial to think about the research question that the researcher wants to address, and then pick the method that offers the most strengths and the fewest weaknesses in addressing that question. If the weaknesses in that approach are difficult to mitigate, an additional method with a different set of strengths and weaknesses should also be applied to the research question; the converging evidence that emerges will make a more compelling case.

References

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech, 34*, 351–366.

Brennan, S. E. & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*, 1482–1493.

Carlson, L. A. & Kenny, R. (2006). Interpreting spatial terms involves simulating interactions. *Psychonomic Bulletin & Review, 13*, 682–688.

Carlson, L. A. & Logan, G. D. (2001). Using spatial relations to select an object. *Memory and Cognition, 29*, 883–892.

Carlson, L. A. (2003). Using spatial language. In B. Ross (Ed.), Psychology of motivation: *Advances in research and theory* (Vol. 43, pp. 127–161). San Diego, CA: Academic Press.

Carlson, L. A., Murray, M. M., Shreiner, C. L., & Carolan, P. Parsing space around objects. (2004). Poster presented at *16th Annual Conference of the American Psychological Society*.

Carlson, L. A., Regier, T., Lopez, W., & Corrigan, B. (2004). Attention unites form and function in spatial language. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-sixth Annual Conference of the Cognitive Science Society* (p. 1536).

Carlson, L. A., West, R., Taylor, H. A., & Herndon, R. W. (2002). Neural coordinates of spatial term use. *Journal of Experimental Psychology: Human Perception and Perception, 2*, 1391–1408.

Carlson, L. A. & Van Deman, S. R. (2004). The space in spatial language. *Journal of Memory and Language, 51*, 418–436.

Carlson-Radvansky, L. A. & Irwin, D. E. (1993). Frames of reference in vision and language: Where is above? *Cognition, 46*, 223–244.

Carlson-Radvansky, L. A. & Irwin, D. E. (1994). Reference frame activation during spatial term assignment. *Journal of Memory and Language, 33*, 646–671.

Carlson-Radvansky, L. A. & Jiang, Y. (1998). Inhibition accompanies reference frame selection. *Psychological Science, 9*, 386–391.

Carlson-Radvansky, L. A. & Radvansky, G. A. (1996). The influence of functional relations on spatial term selection. *Psychological Science, 7*, 56–60.

Carlson-Radvansky, L. A. & Logan, G. D. (1997). The influence of reference frame selection on spatial template construction. *Journal of Memory and Language, 37*, 411–437.

Carlson-Radvansky, L. A. & Tang, Z. (2000). Functional influences on orienting a reference frame. *Memory and Cognition, 28*, 812–820.

Carlson-Radvansky, L. A., Covey, E. S., & Lattanzi, K. L. (1999). "What" effects on "where": Functional influences on spatial relations. *Psychological Science, 10*, 516–521.

Carpenter, P. A. & Just, M. A. (1976). Models of sentence verification and linguistic comprehension. *Psychological Review, 83*, 318–322.

Carpenter, P. A. & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review, 82*, 45–73.

Catlin, J. & Jones, N. K. (1976). Verifying affirmative and negative sentences. *Psychological Review, 83*, 497–501.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*, 30–49.

Clark, H. H. & Chase, W. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Coventry, K. R. & Garrod, S. C. (2004). *Saying, seeing, and acting: The psychological semantics of spatial prepositions*. New York: Psychology Press.

Coventry, K., Prat-Sala, M., & Richards, L. (2001). The interplay between geometry and function in the comprehension of *over, under, above* and *below*. *Journal of Memory and Language, 44*, 376–398.

de Vega, M., Rodrigo, M. J., Ato, M., Dehn, D. M., & Barquero, B. (2002). How nouns and prepositions fit together: An exploration of the semantics of locative sentences. *Discourse Processes, 34*, 117–143.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409–436.

Eilan, N., McCarthy, R., & Brewer, B. (1993). *Spatial representation*. Oxford: Blackwell.

Farah, M. J., Brunn, J. L., Wong, A. B., Wallace, M. A., & Carpenter, P. A. (1990). Frames of reference for allocating attention to space: Evidence from the neglect syndrome. *Neuropsychologia, 28*, 335–347.

Franklin, N., Henkel, L. A., & Zengas, T. (1995). Parsing surrounding space into regions. *Memory and Cognition, 23*, 397–407.

Garnham, A. (1989). A unified theory of meaning of some spatial relational terms. *Cognition, 31*, 45–60.

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9*, 558–565.

Glushko, R. J. & Cooper, L. A. (1978). Spatial comprehension and comparison processes in verification tasks. *Cognitive Psychology, 10*, 391–421.

Guy, R. F. & Norvell, M. (1977). The neutral point on a Likert Scale. *Journal of Psychology, 95*, 199–204.

Hayward, W. G. & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition, 55*, 39–84.

Herskovits, A. (1986). *Language and spatial cognition: An interdisciplinary study of the prepositions of English*. Cambridge: Cambridge University Press.

Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.

Jackendoff, R. (1996). The architecture of the linguistic-spatial interface. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 1–30). Cambridge, MA: MIT Press.

Just, M. A. & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441–480.

Landau, B. & Jackendoff, R. (1993). What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217–265.

Langacker, R. W. (1993). Grammatical traces of some 'invisible' semantic constructs. *Language Sciences, 15*, 323–355.

Langacker, R. W. (2002). A study in unified diversity: English and Mixtec locatives. In N. J. Enfield (Ed.), *Ethnosyntax: Explorations in grammar and culture* (pp. 138–161). Oxford: Oxford University Press.

Lee, J. W., Jones, P. S., Mineyama, Y., & Zhang, X. E. (2002). Cultural differences in responses to a Likert Scale. *Research in Nursing and Health, 25*, 295–306.

Levelt, W. J. M. (1984). Some perceptual limitations on talking about space. In A. J. van Doorn, W. A. van der Grind, & J. J. Koenderink (Eds.), *Limits in perception* (pp. 323–358). Utrecht: VNU Science Press.

Levelt, W. J. M. (1996). Perspective taking and ellipsis in spatial descriptions. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 77–108). Cambridge, MA: MIT Press.

Levinson, S. (1996). Frames of reference and Molyneux's questions: Cross-linguistic evidence. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and Space* (pp. 109–169). Cambridge, MA: MIT Press.

Linde, C. & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language, 51*, 924–939.

Liu, I. (1980). Separating comprehension from the verification process. *Memory and Cognition, 8*, 75–83.

Logan, G. D. (1995). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology, 28*, 103–174.

Logan, G. D. & Sadler, D. D. (1996). A computational analysis of the apprehension of spatial relations. In P. Bloom, M. A. Peterson, L. Nadel, & M. Garrett (Eds.), *Language and space* (pp. 493–529). Cambridge, MA: MIT Press.

MacLeod, C. M., Hunt, E. B., & Mathews, N. N. (1978). Individual differences in the verification of sentence-picture relationships. *Journal of Verbal Learning and Verbal Behavior, 17*, 493–507.

Mainwaring, S., Tversky, B., Ohgishi, M., & Schiano, D. (2003). Descriptions of simple spatial scenes in English and Japanese. *Spatial cognition and computation, 3*, 3–42.

Matell, M. S. & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert Scale items? I. Reliability and validity. *Educational and Psychological Measurement, 31*, 657–674.

Miller, G. A. & Johnson-Laird, P. N. (1976). *Language and Perception* (pp. 374–410). Cambridge, MA: Harvard University Press.

Morrow, D. G. & Clark, H. H. (1988). Interpreting words in spatial descriptions. *Language and Cognitive Processes, 3*, 275–291.

Plumert, J. M., Carswell, C., De Vet, K., & Ihrig, D. (1995). The content and organization of communication about object locations. *Journal of Memory and Language, 34*, 477–485.

Regier, T. & Carlson, L. A. (2001). Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General, 130*, 273–298.

Reichle, E. D., Carpenter, P. A., & Just, M. A. (2000). The neural bases of strategy and skill in sentence-picture verification. *Cognitive Psychology, 40*, 261–295.

Richards, L. V. & Coventry, K. R. (2004). Is it *in* or is it *on*? The influence of geometry and location control on children's descriptions of containment and support events. In L. A. Carlson & E. van der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization and development*. Oxford: Oxford University Press.

Schober, M. F. (1993). Spatial perspective-taking in conversation. *Cognition, 47*, 1–24.

Schober, M. F. (1995). Speakers, addressees, and frames of reference: Whose effort is minimized in conversations about locations? *Discourse Processes, 20*, 219–247.

Seymour, P. H. K. (1973). Stroop interference in naming and verifying spatial locations. *Perception & Psychophysics, 14*, 95–100.

Shoben, E. J. (1978). Theoretical note: Choosing a model of sentence-picture comparisons: A reply to Catlin and Jones. *Psychological Review, 85*, 131–137.

Singer, M. (1977). A constituent comparison model of a picture-first verification task. *Memory & Cognition, 5*, 269–272.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447–481.

Sternberg, S. (1998). Inferring mental operations from reaction-time data: How we compare objects. In D. N. Osherson, D. Scarborough, & S. Sternberg (Eds.), *An Invitation to Cognitive Science*, 2nd ed.: *Methods, models, and conceptual issues, Volume 4*, 365–454.

Talmy, L. (1983). How language structures space. In H. L. Pick & L. P. Acredolo (Eds.), *Spatial Orientation: Theory, research, and application* (pp. 225–282). New York: Plenum Press.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Taylor, H. A. & Tversky, B. (1996). Perspective in spatial descriptions. *Journal of Memory and Language, 35*, 371–391.

Taylor, T. E., Gagné, C. L., & Eagleson, R. (2000). Cognitive constraints in spatial reasoning: Reference frame and reference object selection. *American Association for Artificial Intelligence Technical report, SS-00-04, 168–172*.

Taylor, T. E., Gagné, C. L., & Eagleson, R. (2001). Producing spatial descriptions: Effects of object familiarity. *Proceedings of the 1st International Symposium on Smart Graphics*, 69–75.

Tversky, B. (1975). Pictorial encoding of sentences in sentence-picture comparison. *Quarterly Journal of Experimental Psychology, 10*, 391–421.

Tversky, B. (1991). Spatial mental models. In G. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 109–145). San Diego: Academic Press.

Ullmer-Ehrich, V. (1982). The structure of living space descriptions. In R. J. Jarvella & W. Klein (Eds.), *Speech, place and action* (pp. 219–249). Chichester: Wiley.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation*, Vol. 44 (pp. 35–62).

Zwaan, R. A. & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

Experimental methods for simulation semantics*

Benjamin Bergen

1. Simulation semantics and language understanding

How do people understand language? Though vital to the study of language and the mind, a disproportionately small body of empirical work has historically addressed this question. Research on language understanding presumably falls under the purview of cognitive linguistics, the study of the mind and language, but cognitive linguistics has predominantly produced static, verbal models of linguistic and other conceptual representations, rather than the dynamic models of psychological processing required to explain the processes of language understanding. At the same time, empirical psycholinguistics, whose experimental methods could in principle yield profound insights into the question, has shied away from deep language understanding, preferring to stick to more easily measurable and manipulable aspects of language – principally aspects of linguistic form like syntax and phonology. However, a number of lines of research have recently emerged, which wed psycholinguistic techniques with a cognitive linguistic perspective on language knowledge and use. This work introduces a new, integrated field, which focuses on the idea that language understanding is contingent upon the understander mentally simulating, or imagining, the content of utterances.

This simulation-based view of meaning grows out of theories of language and the mind in which "embodiment" plays a central role. The idea of embodiment in cognitive science (Johnson 1987; Lakoff 1987; Varela et al. 1991; Clark 1997; Lakoff & Johnson 1999; Gibbs 2005) is quite straightforward. It's the notion that aspects of cognition cannot be understood without referring to aspects of the systems they are embedded in – in the biology of the organism, including its brain and the rest of its body, and in its physical and social context. When it comes to understanding language, the embodied perspective suggests that meaning centrally involves the activation of perceptual, motor, social, and

* I owe a great debt of gratitude to Art Glenberg and Michael Spivey for invaluable, formative consultation on experimental methods in cognitive linguistics, as well as to Teenie Matlock, Shweta Narayan, Meylysa Tseng, and Kathryn Wheeler for their useful comments and suggestions on this paper. All remaining errors or omissions are exclusively mine.

affective knowledge that characterizes the content of utterances. The way this works is as follows. Through exposure to language in context, language users learn to pair chunks of language like *kick, Mary,* or *John* with perceptual, motor, social, and affective experiences. In subsequent instances of language use, when the original perceptual, motor, social, and affective stimuli are not contextually present, the experience of them is re-created through the activation of neural structures responsible for experiencing them in the first place. This view of meaning is embodied in that meaning depends on an individual having had experiences in their body in the actual world, where they recreate those experiences in response to linguistic input, and use them to produce meaningful linguistic output.

As for the actual mechanisms underlying language processing on this embodied view, understanding a piece of language is hypothesized to entail performing mental perceptual and motor simulations of its content (Narayanan 1997; Barsalou 1999; Glenberg & Robertson 2000; Bergen et al. 2004; Bergen & Chang 2005). This implies that the meanings of words and of their grammatical configurations are precisely the contributions those linguistic elements make to the construction of mental simulations. The study of how different aspects of language contribute to the construction of mental imagery, and the corresponding theory of linguistic meaning as linguistic specifications of what and how to simulate in response to language, is known as *simulation semantics* (Bergen et al. 2003; Bergen & Chang 2005; Bergen et al. 2004; Feldman & Narayanan 2004; see also foundational work on simulation and language by Bailey 1997; and Narayanan 1997). Beyond language, mental simulation or imagery has long been suggested as a fundamental tool for accessing concepts and their properties (Barsalou 1999; Kosslyn et al. 2001) and recalling events (Wheeler et al. 2000).

This embodied view contrasts with an alternative, disembodied perspective, whereupon understanding language can be entirely characterized as the manipulation of abstract symbols (May 1985). The core proposal of this alternative view is that meanings of words and sentences are like a formal language, composed of abstract, amodal symbols, which stand for aspects of the world (e.g., things, relations, properties – see the discussions in Lakoff 1987 and Barsalou 1999). The substance of mental experience, including meaning, is thus entirely unaffected by its biological, physical, and social context. To understand an utterance, the language user maps words onto the semantic symbols that represent their meanings, and these are then aligned as dictated by the sentence (the symbols for the things take their appropriate positions as arguments for the symbols representing relations). For example, a sentence like *Mary kicked John* would be interpreted by determining which symbolic representations are appropriate for each of the words in the sentence, where *Mary* and *John* are represented by unary symbols, perhaps JOHN and MARY, which contrast with the more complex two-place predicate symbol representing *kicked*, perhaps KICKED(X,Y). These symbols are combined appropriately, so that the meaning of the sentence is something like KICKED(MARY,JOHN). The conceptual system, which as it happens is believed itself to be made up of such abstract, amodal symbols, is consequently updated on the basis of the new information that has just been entered into the system. The content of the utterance is thus understood.

There is very limited empirical evidence for the symbol-manipulation view of language understanding (as argued by Glenberg & Robertson 2000). By contrast, there is substantial support from both behavioral and brain imaging research for the notion that language understanding is based on the unconscious and automatic internal recreation of previous, embodied experiences, using brain structures dedicated to perception and action. This research has begun to uncover the ways in which and the extent to which mental simulation plays a role in language understanding. The goal of the current chapter is to survey the various methods used, with an emphasis on the detailed procedures for performing them. The methods surveyed here address several questions pertaining to mental simulation and natural language understanding: Are mental simulations activated when understanding language? If so, what kind of language triggers them (literal or figurative, concrete or abstract)? What are mental simulations like, and how are they constrained? How are simulations performed, neurally?

Although these questions await conclusive and systematic answers, significant progress has been made, using four principle types of methods. Each is addressed in turn below, with case studies. *Compatibility* effects between the content of language and the performance of actions or perception of percepts tell us that a simulation is performed, and what its motor or perceptual properties are, as do *interference* effects. *Simulation time* effects tell us how the internal dynamics of simulation may be affected by details of the described scenario. Finally, *neural imaging* provides convergent evidence on the localization of simulation.

2. Compatibility effects

Simulation-based theories of language understanding make a straightforward behavioral prediction. If understanding an utterance does indeed involve the activation of perceptual and motor representations, then it should prime these specific, modal representations for subsequent use. For example, it could be that a sentence like *Give Andy the pizza* activates the understander's internal representation of how to move her arm forward (as if to transfer possession of a pizza). If so, then when she is subsequently asked to actually move her real arm forward, she should do so more quickly (the action should be facilitated) than she would without such priming. By contrast, a subsequent action in which the understander performs another, perhaps incompatible action, like moving her arm backwards, should be slowed down (inhibited) by the same sentence. The basic logic of the method is that in order to perform a motor action, one must activate neural motor structures responsible for that particular type of action, and if understanding a sentence leads to increased or decreased excitation of those same neural motor structures, then this should result in quicker compatible (and slower incompatible) actions.

The simulation hypothesis produces the same prediction for perceiving images that it does for performing actions. Just as performing an action should be facilitated or inhibited by preceding language describing compatible or incompatible actions respectively, so perceiving an image (visual, auditory, etc.) should be facilitated or inhibited by language including compatible or incompatible images. For example, a sentence like *The man hammered the nail into the floor* implies a particular direction for the nail, presumably point-down. A very similar sentence, *The man hammered the nail into the wall*, implies a different direction – horizontal – for the nail. We would thus expect to find that tasks involving visual stimuli, where the understander is asked to perceive and perform a categorization task on objects mentioned in a preceding sentence, should show facilitation or inhibition, depending on how well the picture matched the visual characteristics the sentence implied it to have. So after processing a sentence like *The man hammered the nail into the floor*, it should take less time for an understander to perform a task in which she has to perceive a picture of a nail pointing down, and more time for a picture of a nail pointing horizontally.

The main idea that compatibility-type experiments can test is that understanding language activates brain structures responsible for acting and perceiving. They do this through the observation of ways in which language understanding facilitates the performance of described actions or perception of described percepts. Two main lines of research, already hinted at in the above discussion, have addressed this hypothesis, for perception and for action respectively, and each of these is discussed in turn below. We begin with studies demonstrating that after reading a sentence implying an orientation or shape for an object, subjects can perform tasks with an image of that object more quickly if it is displayed with the implied orientation or shape. We then turn to work showing that responses to a sentence are faster when the motion the subject has to perform to respond is compatible with the direction of (hand) motion implied by the sentence.

2.1 Implied object orientation and shape

Does processing sentences automatically and unconsciously activate visual imagery, using visual processing systems? A method for testing this, presented by Stanfield and Zwaan (2001) and Zwaan et al. (2002), had the following setup. The experimenters created pairs of sentences which manipulated the implied orientation (or shape) of objects, like *The man hammered the nail into the floor* versus *The man hammered the nail into the wall*. They then presented an image of the manipulated object, which was either compatible in terms of the manipulated variable, i.e. orientation or shape, or was incompatible. For example, a nail oriented downwards would be compatible with the first sentence above, but incompatible with the second.

Subjects performed one of two tasks. In the first study, subjects were instructed to say as quickly as possible whether the object had been mentioned in the previous sentence. In the subsequent study, they had to simply name the object. This change was motivated by the desire to ensure that any difference in response time resulted from the prior activation of a visual model and was not caused by properties of the task itself. The two tasks differed in that the first, recalling whether the object was mentioned in the previous sentence, explicitly drew the subject's attention to the relation between the sentence and object, while the second, simply naming the object, did not do so at all. Thus, the second method improved over the first since the task it used did not prompt subjects to recall visual memories or even potentially retroactively construct a mental model of the scene described by the sentence.

The research hypothesis in these studies was that the orientation or shape of the object (whichever characteristic was being manipulated) would affect how long it took subjects to respond to images of those objects. This is precisely what was found – when the image of an object was shown with the same orientation or shape it was implied to have in the scenario described by the sentence (e.g. when the nail was described as having been hammered into the floor and was depicted as pointing downwards), it took subjects less time to perform the task than when it was in a different orientation (e.g. horizontal). Zwaan and colleagues also found that when sentences implied that an object would have different shapes (e.g. an eagle in flight versus at rest), subjects once again responded more quickly to images of that object that were coherent with the sentence – having the same shape as they had in the sentence.

In designing a visual compatibility experiment, a number of considerations arise (for more details, see Stanfield & Zwaan 2001; and Zwaan et al. 2002). The first is the nature of the visual property that might be represented in mental simulations of utterance content. While shape and orientation have been studied, others might include color, texture, brightness, size, etc. What is critical in the selection of such a visual feature is that it be possible to construct a large number (perhaps 20–30 pairs) of sentences that, through simple modifications, yield different, incompatible values of the visual property. For example, the sentence *The ranger saw the eagle in the sky* contrasts with *The ranger saw the eagle in the nest* only in the final noun, but the two sentences yield different implied shapes for the eagle (flying versus resting). To ensure that the different sentences can only be reasonably interpreted as implying the predicted visual property, a norming study can be performed, in which subjects (who are different from those participating in the main experiment) are shown the sentence and the two corresponding pictures, and are asked to decide whether one picture, the other, both, or neither matches the sentence.

In selecting or building stimuli, the images that depict the target objects should differ (to the extent possible) only in the relevant visual property. For example, the flying and resting eagle images should be the same size, color, and so on, differing only in the posture the eagle is assuming. A norming study as described in Stanfield and Zwaan (2001) can eliminate the possibility that responses to the pictures are influenced by one version being more canonical or frequent than the other.

One question of potential interest is whether any reaction time differences in the compatible versus incompatible conditions are the result of facilitation, inhibition, or both. In order to assess this possibility, one can use a set of control sentences that mention the target object but do not imply any particular configuration for the visual property in question. For example, *The ranger saw the eagle in the park* does not imply whether the eagle is resting or flying.

Most empirical studies of language processing include filler stimuli, stimuli that are not related to the intent of the experiment, but which are presented to decrease the likelihood that subjects will notice the critical stimuli (the stimuli of interest to the data collection). Experiments like the ones described here often have at least as many fillers as critical sentences. Filler stimuli should be indistinguishable from those in the critical trials. Thus, filler sentences and critical sentences should be the same length, concreteness, etc., and filler images and critical images should be equally large, bright, colored, etc. In half of the trials overall, the image should depict an object mentioned in the sentence (and all of the critical sentence-image pairs should be in this condition). In the other half of the trials (all of which will be fillers), the depicted object should not be mentioned in the sentence. Therefore, depending on the number of fillers included, none to some of them will include an image mentioned in the sentence.

Stimuli should be presented to subjects in four separate lists of stimuli (each subject sees only one of the lists). (If control sentences as mentioned above are included, there is a total of 6 separate lists.) For each critical sentence pair (Sentences 1a and 1b) and its two associated images (Images 1a and 1b), there are four possible presentation combinations: Sentence 1a + Image 1a; Sentence 1a + Image 1b; Sentence 1b + Image 1a; Sentence 1b + Image 1b. Each list should include one of these four versions, for each sentence pair. Thus, if there are 24 critical sentence pairs (resulting in 96 total possible presentation combinations), each list should include a total of 6 of each type of Sentence + Image combination, for a total of 24 critical trials.

Results should be analyzed using Repeated Measures ANOVA (Gravetter & Wallnau 2006; also see Nuñez this volume, for further details on statistical analysis.). ANOVA is a standard statistical test used when independent variables (conditioning factors) are categorical, and when the dependent variable (the thing measured) is continuous. In this particular case, condition (matching versus non-matching (vs. control)) and picture version (which of the two pictures was shown) are within-subjects factors and list (which of the four or six lists was presented) is a between-subjects factor. The hypothesized effect is a main effect of condition, where the matching condition is faster than the non-matching condition.

2.2 The action-sentence compatibility effect

A second compatibility-based method tests the extent to which motor representations are activated for language understanding. The Action-sentence Compatibility Effect (ACE – Glenberg and Kashak 2002) is based on the idea that if language understanders perform motor imagery, using neural structures dedicated to motor control, then understanding sentences about actions should facilitate actually performing compatible motor actions. In ACE experiments, subjects read sentences, of which the critical ones are all meaningful and encode one of two actions – usually motion of the hand away from or towards the body. Subjects indicate whether or not the sentences make sense by pushing a button that requires them to actually perform one of those two actions. For example, in Glenberg and Kashak (2002), sentences encoded movements towards or away from the reader of the sentence, like *Andy handed you the pizza* versus *You handed Andy the pizza.* Subjects started with their hands at an intermediate distance from their body, and then indicated their meaningfulness judgments by pushing a button that was closer to them or farther away from that central location.

The hypothesized effect was an interaction between the direction of motion implied by the sentence and the direction of motion performed by the subject in response to the sentence. Glenberg and Kaschak (2002) report exactly this – a significant interaction between response direction and sentence direction, where responses were faster when the action the subject performed was compatible with the action encoded in the sentence – the Action-sentence Compatibility Effect. For example, responses to sentences that encoded motion toward the subject were faster when the subject had to move her hand towards herself to indicate that it was meaningful than when she had to move her hand away from her body.

A number of considerations are critical to using the ACE to test whether understanding language describing actions makes use of the same cognitive machinery responsible for enacting the same actions. First, the motor actions in question must be both simple to perform and incompatible with each other. Ideally, they use mutually antagonistic muscle groups, like moving one's arm *away* from the body versus moving it *towards* the body, or making a *fist* versus an open *palm* handshape.

Second, these actions should be describable using a broad range of language, and language should exist that can encode one action or the other, depending on a single, simple modification. For example, the verb *catch* implies different handshapes when paired with different direct objects – *catching a marble* involves making a fist while *catching a watermelon* involves more of a palm handshape. Using pairs of sentences that differ only on the basis of a simple modification and strongly imply one type of action or the other decreases the possibility that any ACE effects are produced on the basis of the individual words appearing in the sentences. One way to tell whether sentences imply one type of action or another is to run a norming study, in which subjects are presented with each sentence and asked which of the two actions (or both, or neither) it describes. Moreover, there must be a way for the subjects to respond to the linguistic stimuli (making sensibility or grammaticality judgments) by performing the described action, either by pressing a button placed at a particular location requiring that the subject perform the action to reach it, or simply by performing the action such that it can be videorecorded. The former solution (button press) is preferred to the latter since responses can be automatically recorded, without need for transcription of videotape.

When it comes to filler sentences, there should be at least as many of these as there are critical sentences, and they should be randomly distributed among the presented stimuli, so as to minimize possible effects from trial to trial. As discussed above, filler stimuli should be indistinguishable from those in the critical trials, being of the same length, concreteness, and so on. In half of the trials overall, the sentence should be meaningful (or grammatical, depending on the question subjects are responding to), and all of the critical sentence-image pairs should be in this condition. In the other half of the trials (all of which are fillers), the sentence should not be meaningful (or, in the case of grammaticality judgments, grammatical). Therefore, depending on the number of fillers included, none to some of them will be meaningful. An equal number of the meaningful (or grammatical) and non-meaningful (or ungrammatical) sentences should refer to the same types of action that the critical sentences do, in order to ensure that subjects cannot simply rely on superficial properties of the sentences (thus criteria other than meaningfulness or grammaticality) in order to make their judgments.

Finally, halfway through the experiment, subjects switch which response indicates a 'yes' response and which means 'no'. In order to eliminate order effects, subjects should be randomly assigned to one of two groups, distinguished by the assignment of responses to 'yes' and 'no'.

Results should once again be analyzed using Repeated Measures ANOVA, where sentence-action and response-action (*towards* versus *away*, or *palm* versus *fist*, for example) are within-subjects factors and list (which of the orderings the subject was exposed to) is a between-subjects factor. The hypothesized effect, as mentioned above, is an interaction between sentence-action and response-action, where identical sentence and response actions yield faster reaction times.

2.3 Design issues for compatibility methods

There are several properties of designs for experiments like these that are worth mentioning. First, notice that in designing stimuli for experiments like these, one important consideration is what linguistic sources could potentially yield effects. In the experiments described here, the stimuli never explicitly mentioned the direction of motion, or shape or orientation of the object. Any differences among the sentences, then, must be the product of implied direction, shape, or orientation, which can have resulted only from the construction of a mental simulation of the motor and perceptual content of the utterance. In other words, by making sure that the linguistic stimuli lexically underspecify the independent variable, we can be certain that effects arise not from straightforward lexical meaning, which might itself be a quite interesting effect, but rather from the process of sentence understanding.

Second, these experiments use pairs of stimuli that differ only in a single dimension – the dimension of variation in the implied properties of simulation. Thus, experimenters create sentence pairs like *The man hammered the nail into the floor/wall*, where the only difference is the final word, a word which is not associated with the predicted vertical and horizontal orientation. (In fact, this example is particularly instructive, since wall and floor might in fact in isolation prime the opposite orientation to the one the nail is oriented in – a nail in the floor is vertical, while a floor is predominantly horizontal.) If the main interest of the study is to determine what linguistic properties, e.g. words, phrasal constructions, etc., yield what different effects in visual simulation, then the only differences within pairs of sentences should be the linguistic properties in question.

Third, in their original work on the ACE, Glenberg and Kaschak included only sentences that included the experimental subject in their content, like *Andy handed you the pizza*. However, the simulation semantics hypothesis claims that even understanding language that does not involve the understander, like *Andy handed Sheila the pizza* should engage mental simulation. Indeed, the results from the visual compatibility studies described above, using sentences like *The ranger saw the eagle in the sky* do not involve the experimental subject at all. Subsequent research on the ACE has demonstrated that motor actions described as being performed by someone other than the experimental subject also yield significant effects (Bergen & Wheeler 2005; Tseng & Bergen 2005).

Finally, to position the research described here in terms of the literature, the visual compatibility experiments are a form of priming experiment, in which the presentation of a given stimulus is assumed to yield brain activation that makes faster (or facilitatorily primes) a response governed by brain structures that are identical or connected to those activated by the first stimulus. In particular, in the studies by Zwaan et al. discussed above, the primed response in question is a response to a compatible or incompatible percept. The Action-Sentence-Compatibility effect is a particular sort of priming effect, where the response may itself be compatible or incompatible with the original stimulus. Studies of this form go under the rubric of Stimulus-Response experiments. These effects are analyzed as resulting from "common coding": the neural substrate of the stimulus and the response are overlapping. Other Stimulus-Response Compatibility effects have been shown for spatial location (the Simon Effect – Simon 1969) as well as cross-modal associations like intensity of sound with physical force (Romaiguère et al. 1993; Mattes et al. 2002), among others.

3. Interference effects

Closely related to compatibility effects are *interference* effects. Compatibility effects, as described in the previous section, ostensibly arise from the fact that understanding language about an action activates neural machinery responsible for performing the described action or perceiving the described percept. As a result, identical actions or percepts are facilitated subsequent to language processing. In each of the studies described above, the subjects executed a response after having interpreted the stimulus. As such, these compatibility effects can be seen as a type of priming – a set of neural structures is activated by one activity and thus subsequent use of the same structures is facilitated. Something different happens, however, when the same neural structures are recruited by multiple tasks at the same time. For example, if a particular utterance activates motor or perceptual structures, and if the subject is asked to simultaneously perform another motor or perceptual task, which presumably makes use of the same neural structures, then we will observe not facilitation but interference between the two tasks. To reiterate, interference effects, like compatibility effects, result from the use of the same neural structures to understand language and perform a perception or motion task, but differ from compatibility effects in that understanding the language and performing the perceptual or motor task require the same neural structures to perform different tasks at the same time. The causes for compatibility versus interference effects are somewhat more complex than just temporal overlap, and are discussed in more detail at the end of this section.

Existing interference studies are all based on the use of visual stimuli that are either compatible or incompatible with the presumed simulation evoked by language that is produced at the same time, or immediately before or afterward. However, within this framework, two types of interference have been investigated, deriving from perceptual and motor effects. The first of these lines of research – investigating perceptual interference – is based on an effect known as the Perky effect (Perky 1910; Segal & Gordon 1969). The second is based on recent neuroscientific work on the use of motor systems to perceive and understand actions. Each is addressed in turn below.

3.1 Visual interference effects

Researchers interested in testing whether understanding language with visual content makes use of the visual system look for ways in which processing such language interferes with the simultaneous processing of visual percepts. It has been known for a century that visual imagery can selectively interfere with visual perception. Early work by Perky (1910) had subjects imagine seeing an object (such as a banana or a leaf) while they were looking at a blank screen. At the same time, unbeknownst to them, an actual image of the same object was projected on the screen, starting below the threshold for conscious perception, but with progressively greater and greater definiteness. Perky found that subjects continued to believe that they were still just imagining the stimulus, and failed to recognize that there was actually a real, projected image, even at levels where the projected image was perfectly perceptible to subjects not performing simultaneous imagery.

More recent work on the Perky effect has shown that inteference can arise not just from shared identity of a real and imagined object, but also from shared location. Craver-Lemley and Arterberry (2001) presented subjects with visual stimuli in the upper or lower half of their visual field, while they were performing imagery in the same region where the visual stimulus was, or in a different region, or were performing no imagery at all. They were asked to say whether they saw the visual image or not, and were significantly less accurate at doing so when they were imagining an object (of whatever sort) in the same region than when they were performing no imagery. Performing imagery in a different part of the visual field did not interfere with the visual discrimination task at all.

If Perky effects like these are indeed indicative of visual imagery making use of the same neural resources recruited for actual vision, then they can naturally be extended to language processing. Rather than asking subjects to imagine visual objects, experimenters can ask subjects to process language hypothesized to evoke visual imagery of a particular type – of particular objects with particular properties, or of objects in particular locations. If visual language selectively activates visual imagery, then we should expect a Perky-type effect that results in interference between the displayed visual image and the visual properties implied by the language.

This is precisely the tack taken by Richardson et al. (2003) and subsequently by Bergen et al. (2007). In the work by Richardson and colleagues, subjects heard sentences whose content had implied visual characteristics and then very quickly thereafter performed a visual categorization task where the image either overlapped with the sentence's meaning or did not. They hypothesized that if sentence understanding entailed visual imagery, then there should be Perky-like interference on the object categorization task – it should take longer to categorize an object when it had visual properties similar to the image evoked by the sentence.

More specifically, Richardson et al. suggested that processing language about concrete or abstract motion along different trajectories in the visual field (like vertical versus horizontal) leads language understanders to activate the parts of their visual system used to perceive trajectories with those same orientations. For example, a sentence like *The poacher hunts the deer* implies horizontal motion, while *The ship sinks in the ocean* implies vertical motion. If understanders selectively perform vertical or horizontal visual imagery

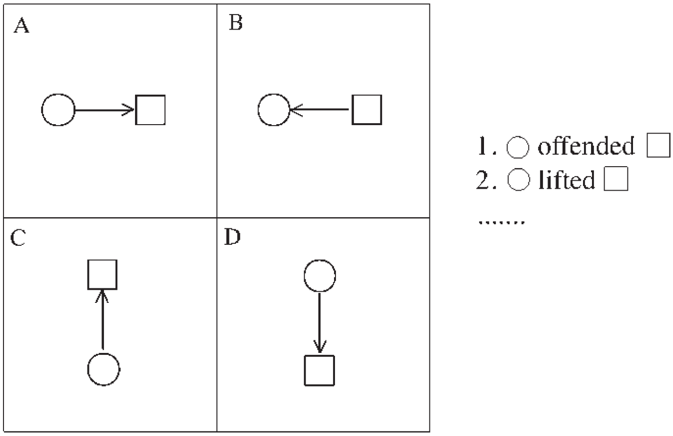


Figure 1. The four possible images from which subjects selected in Richardson et al. (2001)

in processing these sentences, then when they are immediately afterward asked to visually perceive an object that appears in their actual visual field, they should take longer to do so when it appears on the same axis as the motion implied by the sentence. Thus after *The poacher hunts the deer*, subjects should take longer to categorize an object, say as a circle versus a square, when it appears to the right or left of the middle of the screen, but their categorization rate should not be affected when the visual object appears above or below the middle of the screen.

In order to construct stimuli for this experiment, Richardson and colleagues performed two off-line norming studies (described in detail in Richardson et al. 2001), in which subjects provided their intuitions about whether sentences containing particular verbs have vertical or horizontal meanings. The two norming studies used different methods. In the first, subjects simply picked one of four visual images depicting horizontal or vertical motion (depicted in Figure 1) that they decided best captured the meaning of the sentence. In the second, subjects themselves used a graphical interface to produce representations of the scene described by the sentence using the same primitives presented in the images in Figure 1 – circles, squares, and arrows. The interest in both of these tasks was to determine whether language users uniformly assign a vertical or horizontal interpretation to sentences, such that those sentences could be used as stimuli in the Perky experiment.

One additional point of interest here regards the nature of the sentences used. The experimenters were interested in the spatial orientation not just of concrete verbs, like *hunt* and *sink*, but also of abstract verbs, like *respect* and *tempt*. They wanted to determine whether abstract events, just like concrete events, were selectively associated with particular spatial orientations. How abstract concepts are represented and understood is a critical question for all theories of meaning and understanding, but is particularly vital to simulation-based models, because of their reliance on perceptual and motor knowledge. It may not be obvious at first blush what the nature of mental simulations that capture the embodied understandings of abstract notions like *respect* and *tempt* might be. There are insightful discussions of how abstract concepts can be grounded in embodied systems in various places (Lakoff 1987; Barsalou 1999; and Kaschak & Glenberg 2000), so the topic will not be explored in depth here. For current purposes, it will have to suffice that it is an interesting question whether abstract events contain a spatial component. Abstract verbs were thus included in their norming studies, as well as in the Perky experiment described below.

Richardson et al. took verbs, with their associated horizontality/verticality ratings, and presented them to subjects in the interest of ascertaining whether they would induce Perky-like effects on the categorization of visual objects (shapes) that were presented on the screen in locations that overlapped with the sentences' implied orientation. After seeing a fixation cross for 1 second, subjects heard a sentence, then, after a brief pause (randomly selected for each trial from among 50, 100, 150, or 200ms), they saw a visual object that was either a circle or a square, positioned in one of the four quadrants of the screen (right, left, top, or bottom). Their task was to press a button indicating the identity of the object (one button each for 'circle' and 'square') as quickly as possible.

The categorization task the subjects were performing was not transparently related to the sentences that preceded it. This type of design has advantages and disadvantages. One major advantage to having subjects perform a categorization task (rather than a more directly related one, like an up-down or left-right task) is that they are less likely to become aware of the relation between the independent and dependent variables, and thus other high-level cognitive processes involved in reflection are less likely to be activated. In other words, the dependent measure is less likely to be influenced by confounding factors arising from subjects' guessing the purpose of the experiment or simply recognizing the potential relation between the sentence understanding task and the object categorization task.

A disadvantage of this method, compared for example with one in which subjects are asked to imagine the content of the sentences that are presented, is that subjects may be prone to ignoring the content of the utterances, and paying particular attention to the visual objects. In order to eliminate this temptation, it is standard in studies of this sort to include comprehension questions following an unpredictable (at least, to the subjects) subset of the sentences. This ensures that, since they do not know which sentences will be followed by a comprehension question, subjects attend to the meaning of the sentences, even though they don't know that they're really being forced to do so for the purpose of the object categorization task.

Richardson et al. (2003) constructed filler sentences that could be followed by comprehension questions, and added these to the critical sentences, of which fifteen had verbs that had been designated as the most vertical and fifteen the most horizontal, based on the norming studies. It should be noted that in studies in which verbs are more similar to one another (e.g. Bergen et al. 2007), there is a need for a greater number of filler sentences, to obscure the intent of the experiment.

The results were indicative of a clear interference effect – subjects took longer to categorize objects in the vertical axis when they followed vertical sentences than horizontal sentences, and vice versa for objects in the horizontal axis.

A number of subsequent studies have demonstrated Perky effects in language comprehension, and have explored different variations of the effect. The original study did not indicate what parts of the sentences yielded the interference effects that were found. To investigate this question, Bergen et al. (2007) independently varied the verbs and subject nouns in intransitive sentences to determine whether either could single-handedly yield interference effects, and found a strong interference effect when just subject nouns or verbs were associated with a particular region of space. Another methodological modification in that same follow-up study was to split up the vertical axis into the upper and lower regions by using more specifically upwards- or downwards- oriented sentences, rather than conflating them together into a single, vertical condition (a strategy also independently adopted by Lindsay 2003; and Kaschak et al. 2005). Bergen et al. found significant interference within two separate vertical conditions, such that upwards-oriented sentences (like *The mule climbed*) made subjects take longer to categorize objects in the upper quadrant of the screen, but not in the lower quadrant. Finally, the Richardson et al. study did not yield any indication of the role of spatial processing in the comprehension of concrete versus abstract language. The Bergen et al. study separately studied literal (*The mule climbed*), metaphorical (*The rates climbed*), and abstract (*The rates increased*) language, and found that while there were strong Perky effects for literal sentences, there were none in response to abstract or metaphorical utterances.

Another recent follow-up (Lindsay 2003) switched the order of the visual perception and sentence comprehension tasks. In this work, subjects first saw a rectangle move on the screen in one of four directions (upward, downward, rightward, or leftward), and then performed a language comprehension task, in which they read a sentence, and pressed a button as soon as they had understood it. This experiment produced a significant interference effect on reading time of compatible motion – just the same as in the Perky experiments described above, but with the tasks in the reverse order.

Finally, Kaschak et al. (2005) have used a quite similar methodology with slightly different spatial dimensions. In their work, subjects heard a sentence that indicated motion upwards, downwards, toward the subject, or away from the subject, while they simultaneously observed a visual illusion of motion in one of those directions. The subjects' task was to respond as quickly as possible whether the sentence was meaningful or not. The results demonstrated a clear interference effect – it took subjects longer to respond that the sentence was meaningful when it was presented simultaneously with a visual illusion depicting motion in the same direction.

As has been shown above, visual interference effects are reliable and replicable, in a number of methodological permutations. These findings as a whole can be taken as evidence that perceptual systems – in particular the visual system – are unconsciously and automatically engaged in the process of natural language understanding.

3.2 Motor interference effects

As we have seen, the tasks of understanding language about perceptual content and engaging in visual perception can interfere with each other, when performed simultaneously or in rapid succession. Extending this effect to the domain of motion, we can ask whether understanding language about motor actions can similarly interfere with activation of the motor machinery responsible for performing the described actions. To the author's knowledge, no studies currently exist that have tested for interference effects between performing actual actions and understanding language describing the same actions (that is, an interference version of the Action-sentence Compatibility effect, described above). However, a method does exist for testing the activation of motor structures in response to motor language, albeit indirectly, using a type of cross-modal matching.

The use of this cross-modal matching methodology (discussed in Bergen et al. 2003; Bergen et al. 2004; Narayan et al. 2004; and Chan ms.) is based on the relatively recent discovery that perceiving actions activates certain neurons in motor areas that are responsible for enacting those same actions – so-called "mirror neurons" (Gallese et al. 1996; Rizzolatti et al. 1996). Mirror neurons are cells in the motor cortex of monkeys, and presumably also humans, that are selectively activated during the performance of specific motor functions, but which also become active when the individual perceives another person or monkey performing the same function. There are few single neuron studies in humans, but comparable "mirror activity" patterns in humans have been demonstrated through brain imaging (Tettamanti et al. 2005). It has also been established that this mirror system extends to the somatotopic organization of the pre-motor and parietal cortex (Buccino et al. 2001). In particular, the execution or observation of actions produced by the mouth, leg, and hand activate distinct parts of pre-motor cortex, found in ventral sites, dorsal sites, and intermediate foci, respectively. When appropriate target objects are present, there is also activation in a somatotopic activity map in parietal cortex. Mirror neurons and the circuits they participate in have thus been shown to serve dual roles in producing actions and recognizing those actions when performed by others.

Thus, the reasoning goes, when subjects are asked to perceive an action, they activate parts of motor cortex responsible for performing the same action. If they are also asked to simultaneously understand language pertaining to an action, then we may see interference effects when the two actions overlap – just as perceiving an image and simultaneously understanding language that overlaps with that image interfere with each other in the visual domain. Since we know that mirror circuitry is organized by effector, e.g. hand, leg, or mouth, it might be the case that perceiving actions used by a particular effector may selectively interfere with processing language describing actions performed using the same effector. This should contrast with perceiving an action and processing language that indicates actions performed by different effectors, which should not be subject to the same selective interference effect.

In a series of studies (Bergen et al. 2003; Chan, ms.), subjects were shown a stick-figure image of some type of action (performed primarily with the mouth, hand, or leg) and a verb that also described some such action. Subjects were instructed to decide as quickly as possible whether the verb and the image depicted the same action or different actions. The cases of interest were those where the verb and image did not depict the same action. It was hypothesized that subjects should take longer to decide that the verb and image did not match if their actions were both primarily executed using the same effector, compared to the case where they used different effectors. The simulation-based explanation for this hypothesis was straightforward – if perceiving actions and understanding language about actions makes use of motor structures, then very similar actions presented in the two

Match	"hop"	"scratch"	"scream"
Same	"punt"	"juggle"	"lick"
Diff	"juggle"	"punt"	"kick"

Figure 2. Sample stimuli for the image-verb matching task

modalities should yield interference, much like the interference seen in the Perky-type studies described above.

Sixty image stimuli were drawn using a graphics program for these experiments by one of the experimenters, and those images were then paired with verbs through a norming study (described in Bergen et al. 2003). In that study, subjects were asked to provide the word that they thought best captured the action depicted by the image. From the original set of images, 16 images depicting hand, mouth, and leg actions (for a total of 48 images) were selected, on the basis of uniformity of subjects' responses. The matching verb for each image was the one most frequently identified by subjects, and the non-matching verbs using the same and different effectors were randomly selected from among those verbs that no subjects said were the best description of the image. Some of the non-matching pairings were subsequently changed in cases where the experimenters believed the randomly selected verb might plausibly be interpreted as describing the image. In a subsequent study (Chan, ms.), verbs were selected by having subjects choose the best verb for an image from among a small set of options. Some examples of stimuli for the three effectors are seen in Figure 2, along with examples of matching verbs, as well as non-matching verbs in both the same-effector and different-effector conditions:

In the original study (Bergen et al. 2003), the image was presented before the verb. Each trial consisted of a visual stimulus like the images shown above, which was presented for one second, followed immediately by a 500 millisecond interstimulus interval, the first 450 milliseconds of which included a visual mask covering the whole screen. The visual mask was meant to reduce any priming effects that resulted from visual imagery. An English verb in written form was then presented until the subject pressed a button indicating that the verb was or was not a good description of the action depicted in the image. The verb fell into one of the three conditions described above: (1) matching, (2) non-matching same effector, and (3) non-matching different effector.

The results of this experiment were as predicted – subjects took significantly longer to respond that an image and verb did not match when the two actions were produced by the same effector – the mean size of the effect was around 50ms.

One drawback of this method is that it does not eliminate several possible confounds. A first potential confound involves the role of memory in the observed interference effect. It has been shown through neural imaging that recalling motor actions results in the selective activation of parts of motor cortex specialized for performing those same actions (Wheeler et al. 2000; Nyberg et al. 2001). It could be that activation of motor cortex, which produces the interference effect observed in the first study, arises due to demands of the experiment, namely that subjects are required to recall a motor action in order to subsequently decide whether it matches a verb or not. In order to eliminate this possibility, Chan (ms.) reversed the order of the image and verb stimuli, such that subjects were now not recalling an image such that it could be compared to a verb, but were recalling the meaning of the verb, such that it could be compared to the image. In her experiment, Chan found the very same, significant interference effect. She ran the experiment in English and Chinese, and English-speaking subjects took on average 35ms longer to respond in the same effector condition than in the different effector condition, while Chinese-speaking subjects took 85ms longer in the same effector condition.

A second possible confound stems from the possible ambiguity of the images. If the images look in some way like subjects' mental representations of the actions described by the non-matching same effector verbs (for example, if the image of *hop* above looks like it could be an instance of *punt*), then we would anticipate subjects should take longer to reject same-effector verbs than different-effector verbs, simply because they look more like the actions described by the verbs and it thus takes them longer to determine that the verb in fact is not a good description of the action. Notice that this was still a possible confound, despite the experimenters' attempts to pair together same-effector verbs and images that were unambiguous. In order to assess the viability of this explanation, we constructed another version of the task, in which trials paired together not a verb and an image but two verbs (Narayan et al. 2004). In this version of the method, subjects were asked to decide as quickly as possible whether two verbs that were presented in sequence (with just the same procedure used in the previous experiment, except that the image was now replaced by a verb) meant nearly the same thing or not. The same three conditions were possible – the verbs could describe very similar actions (like *run* and *jog*), or could describe different actions, using the same (*run* and *dance*) or different (*run* and *sneeze*) effectors. If the effects we found with image-verb pairs disappeared in this condition, it could be concluded that the original interference effect derived from ambiguity of the original images. If, however, the interference remained in the verb-verb matching study, then properties of the image used would not be a viable explanation for the effect, since no images were used. Narayan et al. once again found an interference effect – subjects took on average 100ms longer to reject the non-matching word pairs when the actions they described used the same effector than when they used different effectors. Thus, the interference effect cannot be due simply to ambiguity of the images used in the original experiment.

The final potential confound requires slightly more explanation. It could be that the difference in response time results from greater overall similarity between the actions in the same-effector condition than between those in the different effector condition. This would mean that effectors might have nothing to do with the effect, which arises strictly on the basis of similarity – namely that it takes longer to reject concepts as identical concepts the more similar they are. Conceptual similarity of actions, regardless of effector identity, is difficult to assess objectively, but a related and quite accessible tool, Latent Semantic Analysis (LSA – Landauer et al. 1998, http://lsa.colorado.edu/) affords a useful substitute.

LSA, among other things, is a statistical method for extracting and representing the similarity between words or texts on the basis of the contexts they do and do not appear in. Two words or texts will be rated as more similar the more alike their distributions are. LSA has been shown to perform quite like humans in a range of behaviors, including synonym and multiple-choice tasks. Of relevance to the current discussion is the pairwise comparison function, which produces a similarity rating from –1 to 1 for any pair of texts. Identical texts have a rating of 1, while completely dissimilar ones would have a rating of –1.

LSA was used to determined the semantic similarity between the presented verb and the verb that was most commonly associated with the particular image in the pretest described above. This similarity rating was then used as a substitute for the conceptual similarity of the actions they denoted. In other words, for the three examples in Figure 1, there was a semantic similarity score between run and run (matching), between run and kick (non-matching, same effector) and between run and drink (non-matching, different effector). This is an indirect way of evaluating the similarity between an image and a verb, since it is mediated by a verb describing the image, but for the time being it has to do in the absence of more direct methodologies.

With LSA ratings assigned to each trial, the average response time per trial (that is, per image-verb pair) was entered into a regression analysis along with the LSA rating for the trial, as described above. This regression included only the non-matching conditions, as including the matching condition (with LSA ratings of 1) produces an abnormal distribution (all matching cases by definition have an LSA value of 1, which is not particularly interesting). Considering only the two non-matching conditions, there was a very weak correlation between LSA rating and reaction time ($R = 0.094$). The trend for subjects to take longer to reject more similar pairs of words and pictures than less similar ones was insignificant ($p = .378$). So while the similarity between a non-matching verb and image as measured by LSA qualitatively seems to account for a small amount of the variance in reaction time, it does not do so significantly. Of course, this does not prove that sharing an effector and not other sorts of similarity is responsible for the reaction time effects we've seen. The LSA rating might be a flawed measure of similarity in general or with respect to verb-image similarity. For this reason, further studies like the ones described below will be required to test whether the effects are actually based on effector identity. The absence of a significant relation between LSA rating and reaction time shown by the regression above does, however, suggest that overall similarity does not transparently account for the interference behavior that was found.

To conclude, cross-modal matching provides a way to test for the activation of motor circuitry during the processing of motor action language. The interference effects shown using this methodology indicate that determining the meaning of an action verb uses overlapping neural resources with the systems responsible for perceiving actions, which themselves are partially constituted by motor structures.

3.3 Interference or compatibility?

As we have seen, interference effects, like compatibility effects, are used to assess access to detailed perceptual and motor representations during various sorts of language processing. But when should one anticipate interference between similar actions or percepts, and when is compatibility most likely? The current state of knowledge is that interference arises when the two matching tasks are performed at the same time, as in the interference studies described above. Kaschak et al. (2005) argue that temporal overlap by itself does not suffice to account for interference effects, but can only do so in combination with integratabilty. On their model, interference effects arise only when the content of a simulation evoked by language is performed simultaneously with a response, and also cannot be integrated with it (to take an example from their study, an image of a spiral apparently moving towards the subject gives the appearance of forward motion but cannot be readily integrated into a simulation of walking towards a building).

Whether or not it is a sufficient condition, it seems that temporal overlap is not a necessary condition for interference. In the image-verb matching task, the stimuli were presented with some delay (500ms), but we nevertheless observe an interference effect. We should not be misled by this case, however – even though the verb and image or two verbs were not presented simultaneously, the interference presumably results from the co-activation of the two motor images required by the task – in order to perform a comparison, both representations need to be activated simultaneously. More critically, though, Lindsay (2003) reports a significant interference effect when the prime and target stimuli were separated by 1500ms. If this finding is reliable, as it appears to be, then it poses problems for the idea that incompatibility effects arise from temporal overlap. One interesting way in which this study differs from the others discussed above that display compatibility effects is in the order of presentation of the sentence and perceptual stimulus. In Lindsay's study, a visual prime (a rectangle moving along the vertical or horizontal axis on a computer screen), preceded a sentence target to be read and understood. By contrast, other such studies with little or no temporal overlap (e.g. Stanfield & Zwaan 2001; Zwaan et al. 2002; Glenberg & Kaschak 2002) all presented the sentence stimulus first. To generalize at this juncture might be premature, but it appears that matching motor or perceptual processes will yield interference when (1) they overlap temporally, <u>or</u> (2) the sentence is presented after the image or action. They will result in compatibility effects when (1) they do not overlap temporally, <u>and</u> (2) the sentence is presented before the image or action.

4. Simulation time effects

Mental simulations are clearly not identical to the real-world experiences they recreate along a number of dimensions, including geometry, physics, and, importantly, time. But it remains an open question precisely how alike real world experiences and simulations are. Since mental simulations appear to include some fine-grained perceptual and motor detail, it might also be the case that they encode some degree of temporal detail – they might unravel over a course of time that correlates positively with the amount of time it would take to perform or perceive the same event in the real world. Events that take longer in the world might take longer to simulate.

In fact, it seems that when people consult a mental image, they scan through it in a way that mirrors actual visual scanning, such that the time it takes to mentally scan from one point in their mental image to another reflects increased time that it would take to actually visually scan between the same two points. Evidence for this observation comes from map tasks. It has been shown that people who study a map and memorize locations on it take longer to mentally scan between landmarks the farther apart they are located on the original map (Kosslyn et al. 1978). More recent work has shown the same effect when a map is simply described, rather than visually inspected (Denis & Cocude 1989).

If imagery time correlates with real time, then simulations evoked by language should take longer, the longer the events they describe take. A straightforward way to investigate the relation between real and simulated time was devised by Matlock (2004). The basic setup of the methodology is to have subjects read a paragraph describing a scenario, which in one way or another evokes relatively slow or relatively fast motion. This paragraph is followed by a final sentence that might be coherent with the paragraph or not – subjects are asked to read this final sentence and decide if it fits with the paragraph. Matlock proposed that language describing slower motion should result in slower, thus longer, mental simulations, and should thus yield longer response times.

As it turns out, Matlock was primarily interested in the processing not of literal motion language, but of fictive motion language. Fictive motion language (Talmy 1996; Langacker 1986) describes static events and scenes using motion language. For example, in *The road meanders through the valley* the road itself does not actually move any more than the fence does in *The fence runs from the house down to the road.* The question Matlock asked was whether the processing of fictive motion sentences displayed time effects, resulting from simulations taking more or less time.

For example, subjects read one of the following paragraphs, intended to evoke motion along a short or a long path:

> *Short Distance Scenario*
> Imagine a desert. From above, the desert looks round. The desert is small. It is only 30 miles in diameter. There is a road in the desert. It is called Road 49. It starts at the north end of the desert. It ends at the south end of the desert. Maria lives in a town on the north end of the desert. Her aunt lives in a town on the south end. Road 49 connects the two towns. Today Maria is driving to her aunt's house. She is driving on Road 49. It takes her only 20 minutes to get to her aunt's house. After she arrives, Maria says, "What a quick drive!"
>
> *Long Distance Scenario*
> Imagine a desert. From above, the desert looks round. The desert is large. It is 400 miles in diameter. There is a road in the desert. It is called Road 49. Road 49 starts at the north end of the desert. Road 49 ends at the south end of the desert. Maria lives in a town on the north end of the desert. Her aunt lives in a town on the south end. Road 49 connects the two towns. Today Maria is driving to her aunt's house. She is driving on Road 49. It takes her over 7 hours to get to her aunt's house. After she arrives, Maria says, "What a long drive!"

Subjects subsequently saw a fictive motion sentence like the following, and were asked to decide whether it related to the story or not:

> *Target sentence:*
> Road 49 crosses the desert.

If subjects took significantly longer to respond 'yes' to the sentence when the paragraph described slow motion than when it described fast motion, then this would imply that subjects were performing longer simulations when the language described motion that would take longer to actually perform or observe.

In order to ensure that any difference in response time did not arise from differences in how well the target sentences fit with the preceding paragraphs, the paragraph-sentence pairs were subjected to a norming procedure. A different set of subjects was asked to rate the paragraph-sentence pairs for how well they went together, on a 1–7 scale. The results showed that the short-distance and long-distance motion paragraphs were indistinguishable in terms of their fit with the final sentence.

The experimental results were exactly in line with the research hypothesis – subjects took 391ms longer to read and make a decision about a fictive motion sentence when it followed a paragraph describing long movement than when it followed a description of short movement. The same significant differences were subsequently replicated for paragraphs that differed in the speed of travel (fast versus slow), and difficulty of the terrain (difficult to navigate versus easy to navigate).

The proposed explanation for these results, that subjects construct a mental model while understanding language and subsequently perform a mental simulation using it to interpret the target fictive motion sentences, is only one of two possibilities. The other is that something about processing language about slow versus fast motion results in slower or faster processing of subsequent language in general. In other words, perhaps the subjects, through processing language about speedy motion or motion over a short path, found themselves in a fast-processing mindset, which resulted in faster responses, regardless of properties of the final sentence. In order to eliminate this possibility, a control study was conducted, in which each of the presented paragraphs was followed by a sentence not encoding motion, fictive or otherwise. Each of these sentences was determined (through another norming procedure) to be comparable in meaning to the fictive motion sentences originally used. For instance, *The road is next to the coast* was a control sentence that was equivalent to the fictive motion sentence *The highway runs along the coast*. If the differences in response time following the short versus long or fast versus slow motion paragraphs had arisen with the non-motion target sentences, then the second explanation, evoking differences in global processing, would be viable. However, no such difference was found

5. Neural imaging

The behavioral evidence from compatibility, interference, and simulation time studies provides strong indications that shared cognitive mechanisms effect the processing of both percepts and actions on the one hand and mental simulations of those same percepts and actions in response to language on the other. However, without convergent evidence from neural imaging studies, it is impossible to draw the strong conclusion that it is those brain areas principally responsible for acting or perceiving that are engaged for language understanding.

The main techniques used for imaging the living human brain are are PET (Positron Emission Tomography) and fMRI (functional Magnetic Resonance Imaging). Both methods are non-invasive and function through the detection of metabolic changes in particular regions of the brain, since increased blood flow correlates positively with neural activity. The two methods detect blood flow to particular regions of the brain in different ways. In PET studies, a radioactive substance emitting positrons is introduced into the subject's bloodstream and blood flow to particular regions is measured by the intensity of positron emissions in those regions. In fMRI studies, nothing is injected into subjects – rather, changes in the magnetic resonance of regions of the brain, resulting from changes in blood flow, are measured using magnetic fields and radio waves. Crucially, though their temporal and spatial acuity differ, both methods allow a snapshot of the brain at a given time, including indications of where neural activity is taking place.

While the phrenologists of the 19th century were mistaken about the possibility of inferring mental properties of individuals from the superficial structure of their skull, they were right that many cognitive functions appear to be at least somewhat localized to particular brain regions, though these may differ more or less among individuals. And neural imaging, along with studies of brain damaged patients, invasive single-cell studies of brain surgery patients, and work with other animals, has provided a great deal of insight into the neural substrates of action and perception. Of particular relevance to the study of simulation, the major responsibility for detailed motor control is shared by a set of motor areas, including primary, supplementary, and secondary motor cortices, as well as regions of the cerebellum. Each of these areas is structured somatotopically – distinct body regions map onto distinct regions of the given brain area, as shown in Figure 3 for primary motor cortex. Similarly, visual cortical areas are arranged retinotopically, such that parts of the retina are mapped spatially onto parts of the visual cortex. During any cognitive behavior, neural activity is not strictly restricted to those brain areas primarily associated with the particular function being performed. Nevertheless, certain brain regions are reliably associated with perceiving objects in particular places in the visual field, and for performing actions with particular effectors.

As it turns out, imagery appears to be selectively executed, at least in part, by the very motor and perceptual areas responsible for the real-world correlates of the particular imagery performed. For example, motor imagery activates the same parts of motor cortex responsible for performing actions using the very same effectors (Porro et al. 1996; Lotze et al. 1999; Ehrsson et al. 2003). Visual imagery selectively activates brain regions responsible for perceiving similar images (Kosslyn et al. 2001), including primary and secondary visual areas. Recalling motor control events also reactivates modal brain structures responsible for performing the very same actions (Wheeler et al. 2000; Nyberg et al. 2001).

Motor and perceptual imagery and memory seem to make use of the specific brain regions used to perceive or perform the given experiences, and we have seen from the behavioral studies above that language understanding makes use of motor and perceptual

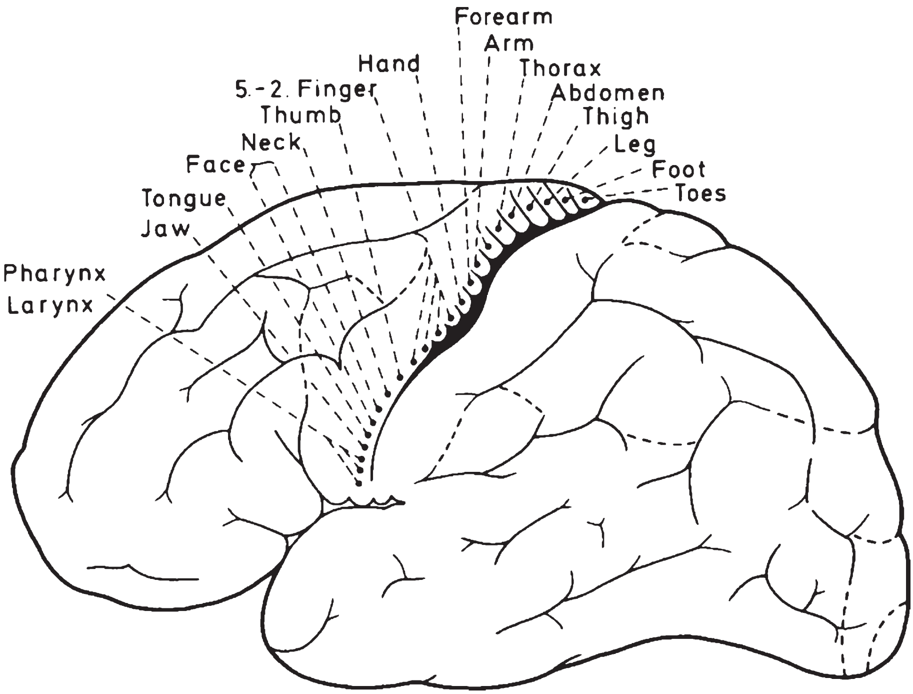


Figure 3. Somatotopic organization of the primary motor strip (from Brodal 1998)

imagery. It would thus be entirely unsurprising for language processing to make use of the same brain regions. Indeed, this is precisely what would be predicted by the simulation semantics hypothesis. Several recent studies show that motor and pre-motor areas associated with specific body parts (i.e. the hand, leg, and mouth) become active in response to motor language referring to those body parts. For example, Pulvermüller et al. (2001) and Hauk et al. (2004) found that verbs associated with different effectors were processed at different rates and in different regions of motor cortex. In particular, when subjects read verbs referring to actions involving the mouth (*chew*), leg (*kick*), or hand (*grab*), the motor areas responsible for mouth versus leg versus hand motion received more activation, respectively. In sentence processing work, Tettamanti et al. (2005) have also shown through imaging that passive listening to sentences describing mouth versus leg versus hand motions activates different parts of pre-motor cortex (as well as other areas).

6. Conclusions

This chapter began by outlining a concrete elaboration of the notion that the content of the mental processes underlying language use are inherently embodied. On this particular view, understanding a piece of language entails performing perceptual or motor simulations, or both. In performing these simulations, the language understander effectively creates or recreates perceptual or motor experiences, using a set of brain structures that overlap with those used to perceive the described percepts or perform the described actions.

The methods we have surveyed here – compatibility effects, interference effects, simulation time effects, and neural imaging, have all incrementally contributed to the body of convergent evidence that now supports the simulation semantics view of language understanding. It appears that language understanders naturally perform both motor and visual simulations, and that the motor and visual systems participate in these processes. Further, they do so in a selective manner – language about events that would by default take place in the upper quadrant of the visual field, for example, specifically makes use of parts of the visual system responsible for perceiving the upper quadrant of the visual field.

Through the application of methods like those described above, and their successors, to a set of progressively more accutely refined questions, the coming years are poised to yield enormous insights into the role of simulation in language processing. Among the major questions that will surely be addressed in detail are the following. What do different types of linguistic units (like different parts of speech, but also different types of phrasal patterns) contribute to the content or form of mental simulations? How do motor and perceptual simulations relate to each other – are they mutually exclusive or can they co-occur (and if the latter, how?) How are different perceptual perspectives taken during the enactment of mental simulations, and how does language trigger these? Are simulations different for speakers of different languages? And finally, how closely do simulations adhere to properties of the real world, like time, space, and physics? With the increasing availability of a broad range of empirical methods, the coming years will bring developments in simulation research that we can hardly imagine.

References

Bailey, D. (1997). *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. Unpublished UC Berkeley Ph.D. Thesis.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–609.

Bergen, B., Narayan, S., & Feldman, J. (2003). Embodied verbal semantics: Evidence from an image-verb matching task. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*.

Bergen, B., Chang, N., & Narayan, S. (2004). Simulated Action in an Embodied Construction Grammar. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*.

Bergen, B. & Chang, N. (2005). Embodied Construction Grammar in Simulation-Based Language Understanding. In J.-O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive grounding and theoretical extensions*. John Benjamins.

Bergen, B., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science, 31*.

Bergen, B. & Wheeler, K. (2005). Sentence Understanding Engages Motor Processes. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*.

Brodal, P. (1998). *The central nervous system: structure and function*. New York: Oxford University Press.

Buccino, G, Binkofski, F., Fink, G., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R., Zilles, K., Rizzolatti, G., & Freund, H. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience, 13*(2), 400–404.

Chan, A. (ms). Embodied verbal semantics in English and Chinese: Evidence from a verb-image matching task.

Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge: MIT Press.

Craver-Lemley, C. & Arterberry, M. (2001). Imagery-induced interference on a visual detection task. *Spatial Vision, 14,* 101–119.

Denis, M. & Cocude, M. (1989). Scanning visual images generated from verbal descriptions. *European Journal of Cognitive Psychology, 1,* 293–307.

Ehrsson, H. H., Geyer, S., & Naito, E. (2003). Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part specific motor representations. *Journal of Neurophysiology, 90*, 3304–3316.

Feldman, J. & Narayanan, S. (2004). Embodied meaning in a neural theory of language. *Brain and Language, 89*(2), 385–392.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain, 119,* 593–609.

Gibbs, R. (2005). *Embodiment in Cognitive Science*. Cambridge University Press.

Glenberg, A. & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review, 9,* 558–565.

Glenberg, A. & Robertson, D. (2000). Symbol Grounding and Meaning: A Comparison of High-Dimensional and Embodied Theories of Meaning. *Journal of Memory and Language, 43,* 379–401.

Gravetter, F. & Wallnau. L. (2006). *Statistics for the Behavioral Sciences* (7th ed.). Wadsworth Publishing.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41*(2), 301–307.

Johnson, M. (1987). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago: University of Chicago Press.

Kaschak, M. & Glenberg, A. (2000). Constructing Meaning: The Role of Affordances and Grammatical. Constructions in Sentence Comprehension. *Journal of Memory and Language, 43,* 508–529.

Kaschak, M. P., Madden, C. J., Therriault, D. J., Yaxley, R. H., Aveyard, M. E., Blanchard, A. A., & Zwaan, R. A. (2005). Perception of motion affects language processing. *Cognition*, *94,* B79–B89.

Kosslyn, S., Ball, T., & Reiser, B. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 47–60.

Kosslyn, S., Ganis, G., & Thompson, W. (2001). Neural foundations of imagery. Nature Reviews, *Neuroscience, 2,* 635–642.

Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago: Chicago University Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Landauer, T., Foltz, P., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259–284.

Langacker, R. W. (1986). Abstract motion. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 455–471.

Lindsay, S. (2003). *Visual priming of language comprehension.* Unpublished University of Sussex Master's Thesis.

Lotze, M., Montoya, P., Erb, M., Hülsmann, E., Flor, H., Klose, U., Birbaumer, N., & Grodd, W. (1999). Activation of cortical and cerebellar motor areas during executed and imagined hand movements: An fMRI study. *Journal of Cognitive Neuroscience, 11*(5), 491–501

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389–1400.

Mattes, S., Leuthold, H., & Ulrich, R. (2002). Stimulus-response compatibility in intensity-force relations. *Quarterly Journal of Experimental Psychology, 55A,* 1175–1191.

May, R. (1985). *Logical Form*. Cambridge: MIT Press.

Narayan, S., Bergen, B., & Weinberg, Z. (2004). Embodied Verbal Semantics: Evidence from a Lexical Matching Task. *Proceedings of the 30th Annual Meeting of the Berkeley Linguistics Society*.

Narayanan, S. (1997). *KARMA: Knowledge-based Action Representations for Metaphor and Aspect*. Unpublished UC Berkeley Ph.D. Thesis.

Nyberg, L., Petersson, K.-M., Nilsson, L.-G., Sandblom, J., Åberg, C., & Ingvar, M. (2001). Reactivation of motor brain areas during explicit memory for actions. *NeuroImage, 14*, 521–528.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21,* 422–452.

Porro C., Francescato, M., Cettolo, V., Diamond, M., Baraldi, P., Zuian, C., Bazzocchi, M., & di Prampero, P. (1996). Primary motor andsensory cortex activation during motor performance and motor imagery: A functional magnetic resonance imaging study. *Journal of Neuroscience, 16,* 7688–7698.

Pulvermüller, F., Haerle, M., & Hummel, F. (2001). Walking or Talking? Behavioral and Neurophysiological Correlates of Action Verb Processing. *Brain and Language, 78,* 143–168.

Richardson, D., Spivey, M., Edelman, S., & Naples, A. (2001). "Language is spatial": Experimental evidence for image schemas of concrete and abstract spatial representations of verbs. *Proceedings of the Twenty-third Annual Meeting of the Cognitive Science Society* (pp. 873–878). Mawhah, NJ: Erlbaum.

Richardson, D., Spivey, M., McRae, K., & Barsalou, L. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science, 27*, 767–780.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Romaiguère, P., Hasbroucq, T. Possamai, C., & Seal, J. (1993). Intensity to force translation: A new effect of stimulus-response compatibility revealed by analysis of response time and electromyographic activity of a prime mover. *Cognitive Brain Research, 1,* 197–201.

Segal, S. & Gordon, P. (1969). The Perky Effect revisited: Blocking of visual signals by imagery. *Perceptual and Motor Skills, 28,* 791–797.

Simon, J. (1969). Reactions towards the source of stimulation. *Journal of Experimental Psychology, 81,* 174–176.

Stanfield, R. & Zwaan, R. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Talmy, L. (1996). Fictive motion in language and "ception". In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 211–276). Cambridge, MA: MIT Press.

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., Fazio, F., Rizzolatti, G., Cappa, S. F., & Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.

Tseng, M. & Bergen, B. (2005). Lexical Processing Drives Motor Simulation. In *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*.

Varela, F., Thompson, E., & Rosch, E. (1991). *The Embodied Mind*. Cambridge, Mass: MIT Press.

Wheeler, M., Petersen, S., & Buckner, R. (2000). Memory's echo: Vivid remembering reactivates sensory specific cortex. *Proceedings of the National Academy of Science of the U.S.A., 97,* 11125–11129.

Zwaan, R., Stanfield, R., & Yaxley, R. (2002). Do language comprehenders routinely represent the shapes of objects? *Psychological Science, 13,* 168–171.

Experimental methods for studying the mental representation of language

Uri Hasson and Rachel Giora

1. Introduction

One of the central assumptions underlying research in cognitive linguistics is that language use reflects conceptual structure, and that therefore the study of language can inform us of the mental structures on which language is based. One of the goals of the field is therefore to properly determine what sorts of mental representations are constructed by various sorts of linguistic utterances. Initial research in the field (e.g., Fauconnier 1994, 1997; Lakoff & Johnson 1980; Langacker 1987) was conducted by way of theoretical discussions, which were based on the methods of introspection and rational reasoning. These methods were used to examine diverse topics such as the mental representation of presupposition, negation, counterfactuals and metaphor, to name a few (cf. Fauconnier 1994).

Unfortunately, the observation of one's mental structures via introspection may be limited in its accuracy (e.g., Nisbett & Wilson 1977). As a result, investigators have come to realize that it is important to examine theoretical claims by using experimental methods (cf., Gibbs 2000; Grady 2000). In this chapter, we will discuss such experimental methods with two goals in mind. Our first goal is to provide the reader who does not specialize in psychology a basic understanding of experimental methods that are used to study the mental representations constructed during language comprehension. Our second goal is to provide examples of how these methods have been used to tackle questions that are of interest to cognitive linguists. Because cognitive linguistic inquiry is interested in the mental representation prompted by language, we restrict ourselves to the mental representations that are later products of the comprehension process. Roughly, these correspond to what is referred to as the interpretation of a certain statement or text. We therefore do not address those initial stages of comprehension, which involve phonological access and access to the mental lexicon. Furthermore, for the sake of brevity, we will mostly restrict the discussion to textual units no larger than a single sentence.

The methods that we will discuss are ones that are often used in psycholinguistic research. These are:

a. Lexical decision and naming measures.
b. Memory measures.
c. Item recognition measures.
d. Reading times.
e. Self report measures.
f. The effects of language comprehension on a subsequent task.

Each of these methods is based on observing an experimental measure to draw conclusions about the mental representations constructed by a certain linguistic unit. For instance, reading times are used to examine the difficulty associated with understanding texts, whereas memory measures look at people's recall to understand how these texts were interpreted. We therefore distinguish between the different research methods based on the dependent variables they focus on (e.g., reading times, proportion of memory errors). In the context of each experimental method we highlight the following:

a. The rationale behind using the dependent variable.
b. Examples of studies using the method.
c. Further considerations regarding the use of the measure, which address advantages, disadvantages and practical problems involved with the use of the method.

2. Lexical decision and naming latencies

2.1 Rationale

In a lexical decision task, participants are presented with a certain letter string on the screen, and their task is to decide (e.g., by pressing a key) whether that letter string makes up a word or not. In a naming task, participants have to read out loud a word presented on the screen. The materials in lexical decision and naming studies typically consist of 50% real words, and 50% non-words (e.g., *flurp*). The rationale behind both tasks is that people should be faster to identify or pronounce a word to the extent that the concept denoted by that word was made accessible by previous context. Response latencies are therefore thought to reflect the mental accessibility of a word. For example, in a seminal study, Meyer and Schvaneveldt (1971) demonstrated that people were faster to decide that *butter* is a word in English when it was preceded by the word *bread* than when it was preceded by the word *nurse*. This kind of facilitation is often referred to as *priming*, and it is attributed to automatic spread of activation between related meanings in the mental lexicon. In the context of lexical decision and naming studies, the first term is often referred to as the *prime*, and the second term is referred to as the *target*. Facilitation is said to occur when response times to a target term are faster after an experimental prime than after an unrelated control prime. Inhibition (or suppression) is said to occur when response times to a target term are slower after the experimental prime than after the control prime.

From the perspective of language comprehension, there are two main factors that are often manipulated in lexical decision and naming tasks. These are (a) the type of prior context, and (b) the interval between the termination of the priming stimulus and the presentation of the target stimulus (sometimes referred to as *interstimulus interval* or ISI). Manipulations of ISI are particularly informative when the research goal is to study the timeline by which meaning is constructed. Note that lexical decision and naming tasks are also useful for investigating which sorts of concepts are made accessible by priming units that are longer than a single word, such as sentences or entire paragraphs.

2.2 Examples

Blasko and Connine (1993) used a lexical decision task to examine the time course of metaphor comprehension. Specifically, they were interested to know how quickly people interpret metaphors, and whether the initial stages of metaphor comprehension involve the construction of a literal-related meaning, as implied by the standard pragmatic model of metaphor comprehension (e.g., Searle 1979). For example, they presented participants with a statement that contained the phrase *hard work is a ladder*, and participants then made a lexical decision either to a literally-related target (*rungs*), a metaphorically-related target (*advance*) or a control target (*pastry*). They found that after participants heard familiar metaphors, response latencies to both metaphorically-related and literally-related targets were facilitated. This finding was interpreted as showing that the comprehension of familiar metaphors does not require the construction of the literal meaning of the utterance before the metaphoric meaning is derived. However, for unfamiliar metaphors the results were different; here only literally-related targets were facilitated in the initial stages of comprehension.

The usefulness of varying the ISI in a lexical decision task can be seen in a study by Till, Mross, and Kintsch (1988), who investigated the time course of sense creation in discourse context. They presented participants with paragraphs that contained statements such as *The servant lit the fire and then prepared the meal. He seasoned it and wrapped it in foil.* The presentation of the paragraph was then immediately interrupted, and participants were given a lexical decision to a critical target word. In one condition, the target was lexically associated with a word that appeared in the sentence (e.g., *tin*), and in another condition, the target word reflected an inference from the statement (e.g., *barbecue*). By manipulating the interval between the termination of the sentence and the presentation of the lexical decision target, the authors found that lexically-associated targets were facilitated fairly early after reading the statements (∼ 200 msec), but inferentially-related targets were facilitated only later on; about 500 msec after reading the statements.

Hasson and Glucksberg (2006) manipulated the interstimulus interval to see what sorts of meaning are prompted by negated utterances. In their study, participants were presented either with affirmative metaphors (e.g., *this lawyer is a shark*) or with their negation (*this lawyer is not a shark*). Participants pressed a spacebar after reading such affirmative or negative metaphors, and were then presented with a target that was related either to the affirmative meaning (e.g., *vicious*) or its negation (e.g., *gentle*). When the target words were presented 100 msec after reading the sentence, then targets related to the affirmative meaning were facilitated after both affirmative and negative metaphors. Targets related to the negative meaning were not facilitated in either case. In contrast, when the target words were presented 1000 msec after reading the sentence, then affirmative-related targets were facilitated after affirmative metaphors, but not after negative metaphors. Again, negative-related targets were not facilitated in either case. The findings show that the comprehension of negative metaphors involves initial activation of the affirmative meaning followed by a reduction of that activation. Giora et al. (2005) conducted a similar investigation of the comprehension of negated adjectival attributions. In their study, participants made lexical decisions to affirmative-related terms after reading either affirmative or negative adjectival attributions (e.g., *this instrument was / was not sharp* followed by the target, *pricking*), and the ISI was a short 100 msec. They found that response latencies to affirmative-related terms were equally fast after reading affirmative or negative adjectival attributions. This finding also suggests that in the initial stages of comprehension, negations are understood as affirmations.

2.3 Further considerations

When used properly, lexical decision and naming methods are useful for studying which sorts of representations are constructed by linguistic units. However, the validity of these measures (like any other measure) depends on the extent to which they reflect the theoretical constructs of interest and are unaffected by other factors. If performance on lexical decision and naming tasks reflected *just* the degree to which a linguistic context affects the accessibility of a certain concept, then these methods would display what is known as high construct validity. In this case, differences in lexical decision times could be unambiguously interpreted as indicating differences in the accessibility of concepts. However, if the performance on these tasks was shown to be susceptible to the influence of other factors, then the results may be of less theoretical interest, because in that case it would be unclear whether task performance actually reflects differences in concept accessibility.

Luckily, the literature on lexical decision and naming tasks has identified a number of exogenous factors that affect performance on these tasks, and so it is possible to circumvent many potential pitfalls (see Neely 1991, for a comprehensive summary of such factors). For example, lexical decision tasks are susceptible to 'expectancy' effects: if participants expect that a certain target will follow a certain prime, and that target fails to appear, then response latencies to the unexpected target will be slowed down. Participants may also wrongly answer 'No' to the lexical decision task when the target does not match the preceding context. Such expectation and backward- checking effects are a result of the decision component of the task and have been reviewed in the literature (see, e.g., Haberlandt 1994; Keenan, Potts, Golding, & Jennings 1990). Another problem is that at long ISI latencies, lexical decision tasks are sensitive to the proportion of pairs in which the prime and the target are related to each other: the greater the proportion of related prime-target pairs in a study, the larger the facilitation seen for the target (McKoon & Ratcliff 1995).

3. Memory measures

3.1 Rationale

The logic underlying the study of language by looking at memory performance is that people's memory for a certain expression could indicate how that expression was mentally represented at the time it was committed to memory, or encoded. In particular, the *errors* that people make in a memory task may be indicative of the representations they constructed. The basic format of memory tasks consists of two stages: a *learning stage*, in which certain expressions are presented to participants, and a *test stage*, which evaluates participants' memory for those expressions. Between the learning and the test stage there is typically a break for a certain period of time, in which participants are engaged in an unrelated task. This break serves to clear working memory from the materials just encountered in the learning stage. The evaluation of memory during the test stage can be carried out by asking people to recall the items presented in the learning stage (free recall), or by presenting them with old and new test items and asking them to indicate for each item whether it had appeared in the learning stage (this is known as an old/new recognition task).

The recognition measure is particularly useful when the goal is to examine the degree to which different expressions are similar in meaning. The greater the similarity in meaning between two expressions, the more likely people are to mistake one for the other in a recognition test. For instance, a hypothesis might be that statement *a* is more similar in meaning to statement *b* than to statement *c*. To test this hypothesis, statement *a* might be presented in the learning stage of a study, and either statements *b* or *c* would be presented in the test stage. Ideally, participants should judge items *b* and *c* as 'new' items, i.e., as items that had not appeared in the learning stage. However, if the results show that item *b* is incorrectly judged as 'old' reliably more often than item *c*, then the results would support the hypothesis in question.

3.2 Examples

The utility of the recognition measure is demonstrated in a classic study by Bransford, Barclay, and Franks (1972). In that study, participants studied sentences, and were later given a recognition task in which they were to asked say whether a certain sentence was presented in the learning stage. The critical materials in the study were ones in which a sentence presented in the learning stage and a sentence presented in the test stage either reflected the same situation in the world, or not. Take, for example, sentences (1) and (2), which describe the same situation in the world:

(1) Three turtles rested on a floating log and a fish swam beneath them.

(2) Three turtles rested on a floating log and a fish swam beneath it.

The authors found that when participants were presented with sentence (1) in the learning stage, they later tended to confuse it with sentence (2) in the test stage; i.e., they incorrectly judged that sentence (2) was presented in the learning stage. In contrast, when the two sentences did not describe the same situation in the world, participants did not tend to confuse them. For example, participants who were presented with sentence (3) in the learning stage seldom mistook sentence (4) for it:

(3) Three turtles rested beside a floating log and a fish swam beneath them.

(4) Three turtles rested beside a floating log and a fish swam beneath it.

These results demonstrated that people are more likely to confuse two statements when the statements refer to the same situation in the world, and indicated that the mental representation of a statement is not *just* a representation of the propositional phrase-structure of the text.

A number of researchers (e.g., Fillenbaum 1966; Just & Carpenter 1976; Smith 1981) have used memory measures to tackle a long-standing question: does the mental representation of negation consist of a representation of what is denied? If comprehending negation involves the representation of the affirmative counterpart, then in the context of a memory task, people who study negated statements should wrongly mistake them for affirmations in the test stage, but the opposite mistake should occur less often. Smith (1981) presented participants with affirmative and negative statements (e.g., *the boy hit the girl* vs. *the boy did not hit the girl*). In the test stage, some of the affirmative statements were presented in negative form, and some of the negative statements were presented in affirmative form. Smith (1981) found that after people learned a negative statement they were more likely to report that they had learned its affirmative form than vice versa. Interesting, this happened for statements with high content of mental imagery (e.g., *the enormous elephant did not lift the fallen tree*), but not for abstract statements (e.g., *his greatest virtue was not his irrepressible confidence*). The fact that abstract negations were remembered better than concrete negations could suggest that the negation of a concrete sentence involved mental imagery of what was said not to be the case. This concrete imagery of the counterfactual state of affairs was later "read off" from memory and mistaken for the representation of an affirmative proposition.

Fillenbaum (1966) examined a related question: is negation always represented as affirmation, or does this depend on the type of concept negated? In Fillenbaum's (1966) study, participants were presented with statements in which adjectives were negated (e.g., the man was not *alive*, the surface was not *rough*). Some negated adjectives offered a direct implication (not alive → dead), whereas some did not (not rough → smooth?). In the test stage, participants were given a forced choice recognition test. For example, after being presented with the sentence *the postman is not alive* in the learning stage, they were asked to recognize which sentence they had seen earlier from the following options: the postman is *alive, not alive*, *dead*, *not dead*. The data of interest were the type of recognition errors that participants made. These fell into two categories:

a. Negation drop: e.g., misrecognizing "not alive" as alive.
b. Gist substitution: e.g., misrecognizing "not alive" as dead, or "not rough" as smooth.

Fillenbaum found that for dichotomous adjectives such as *alive*, participants made more substitution errors than negation-drop errors, but the opposite tendency was found for scalar adjectives such as *rough*. These results suggested that people's encoding of negated propositions may depend on the inference that is afforded by the negation, and that negation is not always encoded as the falsity of the affirmative proposition (cf. Clark & Chase 1972). Fillenbaum's results have been recently extended in two studies using different materials and methods (Mayo, Schul, & Burnstein 2004; Hasson, Simmons, & Todorov 2005).

Sloutsky and Goldvarg (2004) employed memory measures to examine the representation of statements containing logical connectives. Specifically, they were interested to know whether people represent conditional statements of the general form *if p then q* as (more simple) conjunctions (i.e., *p and q*). In the learning stage of their study, they presented participants with statements based on conjunctions, disjunctions, or conditional connectives. These statements were based on abstract contents, e.g., *if he takes medicine, then he likes the zoo*. In the following test stage, they presented participants with old and new statements. The crucial statements in the recognition stage were new statements that differed from an original sentence only in the logical connective. They found that after studying a conditional statement, participants were more likely to confuse that conditional with a conjunction form (e.g., *he takes medicine and he likes the zoo*) than to confuse the conditional with a disjunction form (*he takes medicine or he likes the zoo*). The findings suggested that people might build a 'minimal' representation of conditional statements: although a conditional is logically consistent with three sorts of states of affairs, people seem to represent that possibility in which both the antecedent and consequent are true, and therefore later confuse between conjunctions and disjunctions.

3.3 Further considerations

Though memory measures often provide useful information for evaluating certain hypotheses, there are certain methodological weaknesses that are inherent in such methods. As a result, it is best to use such measures in combination with other methods targeting the questions of interest (this is generally true).

One weakness is that memory performance on a recall or recognition task does not depend solely on the way in which a certain linguistic element was encoded during the learning stage, but also on various processes that occur during the memory test stage itself. This sort of weakness is inherent in various measures that are conducted "offline", that is, after participants had finished comprehending the stimuli. Consider for example Smith's (1981) finding that people are more likely to mis-recall negation as affirmation than vice versa. This finding was interpreted as showing that negation is sometimes represented and encoded as affirmation. However, the finding might also be a result of a certain response bias that is manifested in the test stage: when participants are unsure whether a sentence was presented in affirmative or negative form, they might answer, for whatever reason, that the statement was originally presented in affirmative form.

Also, the instructions given to participants in the learning stage may affect how participants understand the statements, and consequently, their performance in the test stage. In particular, participants may use specific comprehension and encoding strategies when they are told to memorize the materials presented in the learning stage, and these strategies may be very different from those used in routine comprehension (cf. Clark & Clark 1977: Ch. 2).

Finally, one practical consideration is that materials presented at the beginning and end of the learning stage are often recalled better than items in the middle of the list (Brown 1958; Peterson & Peterson 1959). For this reason, it is recommended that the first items and last items presented in the learning stage be "filler items" – i.e., materials that will not be analyzed. Also, it is advisable to randomize the order of presentation of the items of interest in both the learning and test stages.

4. Item recognition measures

4.1 Rationale

This method, like lexical decision and naming tasks, seeks to establish the relative accessibility of a certain term following text comprehension. In practice, participants read a text, and immediately after reading that text (typically, within 1 or 2 seconds), they are asked to decide whether a term had appeared in the text they had just read. The latency to make the decision is taken to reflect the relative accessibility of the term. More generally, it is assumed that different levels of accessibility reflect levels of activation for the concept referred to by the term.

4.2 Examples

MacDonald and Just (1989) used the recognition method to examine whether the representation of negation differs from affirmation. Participants were presented with statements that referred to two entities, one of which was negated (e.g., Elizabeth baked *some bread*, but *no cookies*). Participants read such statements at their own pace, and after each statement they were presented with a term on the screen, and had to verify whether it had appeared in the statement or not. In the crucial trials, the terms presented for verification corresponded either to the negated term (e.g., *cookies*) or to the non-negated term (e.g., *bread*) in the sentence just read. MacDonald and Just (1989) found that verification times were slower for negated terms than for non-negated terms suggesting that, "negation decreases accessibility of a negated noun" (p. 641).

The same procedure was used by Kaup (2001) to examine a different hypothesis: that the mental representation of negation depends on whether the negated term refers to an entity that is present or absent from the situation described in the sentence. Kaup (2001) presented participants with statements such as (5) and (6):

(5) Almost every weekend, Mary bakes *some bread* but *no cookies* for the children.

(6) Elizabeth tidied up her drawers. She burned the old *letters* but *not the photographs*.

Note that in (5), the negation implies absence of cookies from the scene, whereas in (6), the negated term is implied as present while the affirmative one is absent. The results showed that, on the whole, verification latencies were slower for negated terms than for affirmative terms. However, this difference between verification latencies for the negated and non-negated terms depended on whether the negated terms referred to items absent or present in the scene. The difference between response latencies to negated and non-negated nouns was largest when the negated term was absent from the scene and the non-negated term was present in the scene (example 5). When the negated term was present in the scene and the non-negated noun absent (example 6), then the difference between verification latencies for affirmative and negated terms was smaller. Such findings indicate that negation reduces the accessibility of terms in its scope, but they also highlight the importance of a second factor, viz., the presence or absence of a given entity in the situation described.

The item verification method was also used in a study conducted by Glenberg, Meyer, and Lindem (1987), which examined whether elements that are foregrounded in a text are more cognitively accessible. In their study, participants read a story in which a certain element was either foregrounded or not. For example, one story described a person preparing for a marathon, who *takes off* his shirt before jogging around the lake. A different version of this story was constructed, in which a person is described as *putting on* the sweatshirt before jogging around the lake. After reading the story, participants were asked whether a certain word, e.g., *sweatshirt*, had appeared in the story. The authors found that (in certain conditions), people were faster to verify the appearance of a term in the text when that term corresponded to a foregrounded entity than when the term corresponded to a non-foregrounded entity. They argued that these findings support the notion that people build mental-model representations of the discourse they are reading – a representation of what the text is about (for alternative explanations, see McKoon & Ratcliff 1992).

A study by Horton and Rapp (2003) makes a similar point. They presented participants with stories in which a certain object was implied to be occluded from the protagonist, or not. For example, the protagonist might be described as observing a vase, and later on, a certain action was performed that may or may not have resulted in the occlusion of the vase from the protagonist. After reading these stories, participants were asked whether a vase was mentioned in the scene just described. Participants were slower to verify that the object had appeared in the scene when the object was described as being occluded from the viewpoint of the protagonist than when the object was not implied to be occluded.

Finally, item verification may also be used to study inferences prompted by the text. In such studies, researchers may be interested in the speed by which people accurately determine that a certain word did not appear in the preceding text. For example, people might take longer to decide that the word 'hammer' did not appear in sentence (7) than to decide that it did not appear in (8), because the term is implied more strongly by the former:

(7) He pounded the nails into the wood.

(8) He bought some nails and wood.

4.3 Further considerations

As reviewed, item verification measures are useful for examining hypotheses concerning the degree of accessibility of different terms appearing in a text. Their relative drawback is that it is sometimes unclear whether differences in verification latencies necessarily indicate differences in the mental activation of concepts. Take, for example, the findings by MacDonald and Just (1989) and Kaup (2001), which showed that verification latencies for a term were longer when it was in the scope of negation. These findings were interpreted as showing that negation reduces the activation level of concepts. However, the increased verification latencies are also consistent with the possibility that negation prompts the construction of two mental spaces: a factual and a counterfactual one, and that this more complex representation of negated sentences leads to longer verification latencies for the negated items, because those items are represented in two spaces. It is therefore useful to corroborate the results of such studies by other measures of accessibility, such as lexical decision tasks.

Item recognition, like lexical decisions, involves a decision stage which is non-automatic. Because a decision is involved, the responses on such tasks could be biased by backward-checking heuristics of the sorts that also affect lexical decision tasks. For example, participants may check for the compatibility of the target item with the previous context and this could affect verification times. It is therefore suggested that there be a short latency between the termination of the sentence and the presentation of the target term (McKoon & Ratcliff 1986) to minimize strategic processes.

5. Reading times

5.1 Rationale

In some experiments, the time needed to read a text is taken to be indicative of the processing difficulty that is associated with the comprehension of that text. Reading times are particularly useful for examining whether preceding contexts have differential effects on text comprehension. In such designs, a given text appears in different experimental conditions that vary in the type of context that precedes the text. Reading times are also used to compare the processing difficulty of different expressions, but in such cases it may be more difficult to interpret any differences between experimental conditions.

5.2 Examples

Gentner, Bowdle, Wolff, and Boronat (2001) report a study that examined whether the comprehension of metaphorical statements activates the source domain of the metaphor. They constructed two sorts of experimental conditions. Both conditions ended with the exact same statement, e.g., a statement drawn from the mapping A DEBATE IS A RACE. In one condition (the consistent condition), the preceding statements in the paragraph were instances of the same cross-domain mapping. In the other condition (the inconsistent condition), previous statements were instances of a different mapping, e.g., A DEBATE IS A WAR. The authors found that the last statement in the paragraph was read faster in the consistent condition than in the inconsistent condition (see also, Allbritton, McKoon, & Gerrig 1995). These findings suggest that when people read metaphorical statements, they activate rich conceptual structures that correspond to the mappings between the source and target domains, and the accessibility of this knowledge assists the comprehension of later statements based on the same mapping.

Unfortunately, differences in reading times are sometimes more difficult to interpret. Johnson (1996) examined comprehension times for metaphorical statements and similes. Johnson argued that if metaphors are simply disguised similes, then metaphors should be understood just as fast as similes, or somewhat slower than similes. Participants read similes and metaphors at their own pace, and it was found that similes were read more slowly than metaphors (a control condition demonstrated that this was not due to similes containing an additional word). Johnson (1996) interpreted the findings as suggesting that similes are mentally transformed into class inclusion statements and therefore take longer to comprehend. However, the data are also consistent with the interpretation that metaphors and similes are understood by different mental processes which lead to distinct mental representations. Comprehension latencies alone cannot decide between these interpretations.

In a more elaborate study of this topic, Gentner and Bowdle (2001) report an experiment in which they manipulated the sort of source domain that appeared in the metaphors and similes so that in some cases the source was novel (e.g., a novel is / is like a *glacier*), whereas in other cases the source had a conventionalized sense (e.g., a gene is / is like a *blueprint*). They found that when sources were novel, then similes were comprehended faster than metaphors, but when sources were conventionalized, then metaphors were comprehended faster than similes. They interpreted the findings as showing that when sources are novel, both tropes are understood as comparisons. Therefore similes have an advantage, as they indicate the comparison directly. But why were opposite results found for conventionalized sources? Gentner and Bowdle suggested that (a) when sources have a conventional meaning, then metaphors are understood as categorization statements, whereas similes are understood as comparisons, and (b) that categorization statements are inherently easier to comprehend than comparisons. As can be seen from this discussion, reading times alone are weak constraints on theoretical accounts. Therefore, such data may not be sufficiently informative if the purpose is to make specific claims about the online construction of mental representation. In such cases they are useful when considered alongside other sorts of evidence.

Reading times are more useful when there is an a priori hypothesis that one sort of statement is more difficult to process than another. McElree and colleagues (McElree, Traxler, Pickering, Seely, & Jackendoff 2001; Traxler, Pickering, & McElree 2002) have examined whether sentential contexts that are thought to involve type-shifting (9) take longer to understand than those that do not demand shifting (10).

(9) The author was starting the book in his house on the island.

(10) The author was writing the book in his house on the island.

Verbs like "starting" typically require an activity as a complement, but in certain cases they can be followed by a noun. In such cases, the noun phrase has to be interpreted as referring to an action (e.g., writing the book). In this study, the sentences were presented one word at a time, in a self-paced manner, and each statement was followed by a comprehension question to ensure that participants read for comprehension. The authors found that the noun (e.g., *book*) was read more slowly in they type-shifting context (9) than in the non-shift context (10), and the same held for the word following that noun. These results are consistent with the idea that type-shifting contexts involve further psychological elaboration of the noun phrase.

Note that there are alternative explanations for these data. For instance, the phrase "writing the" could result in an expectation for the word "book", whereas the phrase "starting the" would not. If so, the findings would not reflect type shifting, but differential predictability of the critical noun in the two sentential contexts. For this reason, the authors took great care to eliminate such alternative explanations for the data. For instance, they verified that the sentences that were used in the different conditions were equally *plausible*. They also verified that the verbs used in these sentences were equally *frequent* in use. Finally, they verified that the noun phrase (e.g., *the book*) was not better predicted by one verb than the other.

Black, Turner, and Bower (1979) observed reading times to see whether readers are aware of points of view established in narratives. They constructed statements that were based on deictic verbs, so that in some cases a consistent perspective was maintained throughout the sentence (e.g., 11), and in other cases there was a change in perspective in mid sentence (e.g., 12).

(11) Bill was sitting in the living room reading the paper when John *came* into the living room.

(12) Bill was sitting in the living room reading the paper when John *went* into the living room.

Participants took longer to read those sentences that involved a change in perspective. These results were interpreted as showing that readers are sensitive to narrative perspective and prefer consistent viewpoints.

5.3 Further considerations

Reading-time measures are a relatively precise measure, and one that is easily implemented in a lab. Depending on the goal of the study, an experimenter might want to present the materials one sentence at a time or one word at a time on the computer screen. The rate of presentation may be predetermined by the experimenter, or self-paced by the participants in the study. Presenting the sentences one word at a time can help identify points of difficulty in the reading of the sentence, but has the drawback of making the task less natural than normal reading. In particular, when the words are presented one at a time, readers cannot go back to parts of the sentence they had already read, and cannot look ahead at upcoming words – both of which are possible in normal reading. These might be some of the reasons for why reading times for individual words are typically much longer than when those words appear in the context of a sentence (see Haberlandt 1994, for a review).

6. Self report measures: Listing features, choosing features

6.1 Rationale

In a feature-listing task, participants are typically asked to write down properties that best capture the meaning of a certain expression. For instance, they might be asked to write down (or choose from a list) which properties are implied by a metaphorical statement such as *this lawyer is a shark*. In another variant of the task, they might be asked to choose from a list which property (or paraphrase) best captures the meaning of the expression. The rationale behind the method is that if people can understand the meaning of an expression, then they can also explain what that expression means and which features are associated with it. This method is often used to examine how linguistic contexts affect sense generation.

6.2 Examples

Interesting examples of the use of features-listing are found in studies of noun-noun combinations. Typically, such studies aim to examine the relation between the meaning of the constituents of the combination and the meaning of the combination itself. For instance, Wilkenfeld and Ward (2001) examined to what extent the properties of a noun-noun combination (e.g., *motorcycle-carpet*) reflect the meaning of the constituents of that combination. In their study, one group of participants wrote down salient features of the individual constituent nouns (e.g., *motorcycle*, *carpet*), whereas a different group of participants defined the combination, and then wrote which features were associated with the definition. For instance, one definition of *motorcycle-carpet* was "a field of thousands of motorcycles", and a feature listed for this definition was *crowded*. This design allowed the authors to examine the extent to which features of the individual constituents overlap with those properties associated with the combination. They were particularly interested to see whether combinations can imply features that are not directly reflected in their constituents – i.e., emergent features. The authors operationalized an emergent feature as one that was listed for a combination, but not for either of the individual constituents alone. They found that when the terms in the combination were dissimilar (e.g., as in the combination *couch-skate*) then 28% of the features listed were emergent. When the terms were similar (as in *zebra-horse*), 21% of the features were emergent. The analysis in this study was based simply on whether or not a feature of the combination was mentioned for its constituents.

A more detailed analysis of this issue was undertaken by Hampton (1987), who examined whether important features of the constituents necessarily end up being important features of the combination. In one experiment, participants either listed features for individual terms (e.g., *sports*, *games*) or for conjunctive combinations of those terms (e.g., *sports that are also games*). This study revealed that some of the combination features were emergent; i.e., were not listed for either of the constituents. In a second study, Hampton asked participants to rank order the importance of the features for the definitions of concepts. The ratings were made for individual terms and for combinations of the terms.

Features that were rated as important for constituents were usually rated as important for the conjunctions of those concepts. However, there were a few cases (less than 10%) in which features that were unimportant for the constituents tended to be rated as important for their conjunction. For example, the properties *small* and *lives in a cage* were rated as important for the conjunction *pets that are also birds*, but as unimportant for *pets* or *birds* separately.

Johnson and Keil (2000) also employed a feature-listing task to study the interpretation of noun-noun combinations. They used combinations that were interpretable but not completely familiar (e.g., *hospital-rat*, or *mountain-knife*). In their study, they asked participants to write thirty-six properties characteristic of each noun, and five properties characteristic of each combination. From this corpus, they constructed two master lists: one that contained properties listed for the nouns, and one that contained properties listed for the combinations. Note that comparing these two lists is sufficient for evaluating the prominence of the noun-features in the combination. However, before they analyzed the overlap between the lists, Johnson and Keil (2000) first derived two reduced lists from these lists. These reduced lists included only those properties that were mentioned by at least half of the participants for each of the nouns or for each of the combinations. These reduced lists therefore captured features that may be considered as more typical. They found that only 32% of the typical features of the combinations were also listed as typical features of the head noun. Thirty percent appeared as non typical features of the head noun, and 38% were not mentioned as features of the head noun (e.g., the property *stationary* was listed for the combination *hospital-bicycle* but not for *bicycle*).

Feature listing can also be used to examine how different sentential contexts mediate the accessibility of noun-features. Coulson and Matlock (2001) examined the meanings associated with words that appeared in different sentential contexts. In the null context, a term (e.g., *anchor*) was presented alone, and participants listed properties associated with the term. In addition, there were three experimental conditions, in which the terms were embedded in different sentential contexts. These contexts made up either a literal context (e.g., *he almost forgot about the anchor*), a metaphorical context (*his wife was his anchor*), or a literal-mapping context (*we were able to use a barbell for an anchor*). When appearing in these sentences, the terms of interest appeared as the terminal words of the sentence. Participants read each sentence and then quickly listed two or three features for the term of interest, which was underlined. Coulson and Matlock (2001) found that, on the whole, a significant proportion of the features listed in the sentential contexts (~40%) was not listed for these terms in the null context condition. The authors then went on to analyze which set of sentential features was most *similar in meaning* to those given in the null context condition. To this end they used a *latent semantic analysis* method (Landauer, Foltz, & Laham 1998). They found that the features given in the metaphorical context were the least similar in meaning to those given in the null context condition, and that the features elicited in the literal context were most similar in meaning to that of the null context. A feature-listing task of this sort was also used to study if interpretation of metaphorical statements generates emergent features (Becker 1997).

In another variant of this task, participants might be asked to choose which properties or paraphrase best capture the meaning of an expression from a list of pre-constructed options. For example, Costello and Keane (2001) examined noun-noun combinations to determine which properties of the modifier are attributed to the head noun (e.g., *moth*) in combinations such as *bumblebee-moth*. They examined four sorts of possible properties, which varied on two dimensions: whether the property was related to a common dimension of the categories, and whether the property was diagnostic of the modifier. For example, for the combination *bumblebee-moth* these four options were:

a. A moth that is black and yellow (this property is related to a common dimension; color, and color is diagnostic of bees).
b. A moth that is the size of a bumblebee (this property is related to a common dimension; size, and size is non-diagnostic of bees).
c. A moth that stings (this property is unrelated to a common dimension, but having a sting is diagnostic of bees).
d. A moth that fertilizes plants (this property is unrelated to a common dimension, and fertilizing plants is non-diagnostic of bees).

They found that participants preferred interpretations that were based on diagnostic properties, as in example *c* above (68% of the interpretations). Whether or not the property was related to a common dimension had no effect on participants' choices. The same results were found when participants wrote down their definitions, and these definitions were then categorized into the four sorts of interpretations outlined above.

6.3 Further considerations

Feature-listing studies are easy to conduct, and often provide intriguing data. However, their downside is that they are based on probing people's conscious impressions of the meaning of an expression, which might not accurately reflect mental representation. Furthermore, by the time that people list the features of an expression, its initial interpretation may no longer be accessible. In fact, there is no way of knowing whether the features that people list capture the direct meaning of the expression, or people's implicit theories about the meaning of such expressions. Take for example the study by Costello and Keane (2001). The authors assumed that people's choices in that task reflect direct and unmediated access to the interpretation of the expression. However, people might hold lay theories about what meanings are associated with such expressions (theories that might be quite similar to those suggested by the authors!), therefore obtaining the results found in the study.

7. The effects of comprehension on subsequent tasks

7.1 Rationale

In certain cases, it is possible to examine how people interpret language by observing aspects of their behavior in a subsequent context. The rationale behind the method is that to the extent that a linguistic expression invokes a mental representation, that mental representation can be studied by observing the effect it has on subsequent behavior.

These sorts of paradigms are often called "priming" paradigms. Often such designs are employed to examine how the comprehension of one linguistic expression affects the processing of another expression (for example, whether the reading of the word *doctor* speeds up the subsequent reading of the word *nurse,* Meyer & Schvaneveldt 1971). We have discussed such cases in our discussion of lexical decision and naming tasks. In other cases, a researcher might be interested to find out how comprehension affects non-linguistic behavior. The advantage of studying the effects of comprehension on non-linguistic behavior lies in that participants are unaware that their input may reflect a product language comprehension, and so there is less risk that the results are due to strategic thinking on behalf of participants. Furthermore, such methods are sometimes the only way to answer specific research questions. This section focuses on such cases.

7.2 Examples

Estes (2003) used an indirect measure of comprehension to study how people understand two sorts of noun-noun combinations. Some theories (e.g., Gagne 2001) argue that the comprehension of noun-noun combinations is based on finding the relation that links the modifier and the head. It is assumed that there exist many sorts of such relations; for example, the relation "is like" could mediate the comprehension of the *attributive* combination *cactus-carpet*, where a property of cactus is attributed to carpet. A different relation would mediate the comprehension of the *relational* combination *pancake-spatula*. Arguing against this view, Estes (2003) suggested that when people understand attributive combinations, they engage in comprehension procedures that are qualitatively distinct from those used to comprehend relational combinations. In his study, he first asked participants to define (i.e., write the meanings of) attributive and relational combinations. Participants were then asked to judge the similarity of the terms in the combination; e.g., "how similar are cacti and carpets?". These similarity ratings were compared to the ratings given by a control group that had not defined the combinations beforehand. The results showed that after participants defined attributive combinations, the terms in the combination seemed less similar than they seemed in the control condition. However, after defining relational combinations, the similarity of those terms was rated as greater than it had been in the control condition. The results suggested that the meaning of attributive combination might be arrived at via a comparison process, which highlights differences, whereas relational combinations are understood in a different manner.

The word fragment completion task is another task that can reveal which sorts of meanings are accessible. In this task, participants are typically presented with certain materials in the first stage of a study, and are later asked to complete a fragmented word (b-tt-r) with the first word they can think of (*bitter*, *butter*, *better*, *bettor*). A number of studies using this task have shown that the semantic processing employed in the first stage of the study, e.g., reading behavioral descriptions or a short story, results in indirect priming that is evident in people's word completions (e.g., Richards & French 1991; Whitney, Waring, & Zingmark 1992).

For instance, Giora and Fein (1999) presented participants with short stories that biased a literal or ironic interpretation of the last sentence in the story. After reading these vignettes, participants were asked to complete word-fragments related to either the literal or the ironic meaning of the targets. For example, the sentence *Moshe, I think you should eat something*, was embedded as the terminal sentence in either an ironic context (13a) or literal context (13b):

(13) a. After he had finished eating pizza, falafel, ice-cream, wafers and half of the cream cake his mother had baked for his brother Benjamin's birthday party, Moshe started eating coated peanuts. His mother said to him: "Moshe, I think you should eat something".
b. At two o'clock in the afternoon, Moshe started doing his homework and getting prepared for his Bible test. When his mother came home from work at eight p.m., Moshe was still seated at his desk, looking pale. His mother said to him: "Moshe, I think you should eat something".

Findings showed that ironically biased contexts (13a) facilitated both contextually compatible ironic responses as well as contextually incompatible-but-salient responses (related to the literal meaning of the irony). In contrast, literally biased targets (13b) facilitated only the salient, literally related concepts. Such findings showed that salient meanings were retained even when incompatible.

Another variant of this method is one where researchers examine how being exposed to a certain stimuli later affects people's behavior on a task that is ostensibly completely unrelated. This variant capitalizes on the fact that people's recent experiences affect how they comprehend ambiguous or neutral stimuli. In a classic study, Higgins, Rholes, and Jones (1977) presented participants with a description of a person, whose behavior was judged by participants in a control group to be mid-way between adventurous and reckless on a bipolar scale. Two other groups of participants also judged the behavior of the character on the basis of the same description, but only after being earlier exposed to certain words in the guise of an ostensibly unrelated 'perception study'. One of these groups was exposed to words related to recklessness, whereas the other had been exposed to words related to adventurousness. Higgins, Rholes, and Jones (1977) found that those participants that had been exposed to words related to recklessness later judged the character to be more reckless than adventurous, whereas participants that were exposed to words related to adventurousness later judged the character as more adventurous than reckless. These findings indicated that when people try to interpret an ambiguous meaning, they may be affected by cognitive constructs that have recently been made accessible.

McGlone and Harding (1998) used the disambiguation rationale to study whether people represent ego-moving and time-moving perspectives when comprehending temporal expressions. They presented participants with texts that contained ego-moving expressions (e.g., *we will arrive at the exam date in two days*) or time-moving expressions (e.g., *the exam date will arrive in two days*). Following, they asked the participants to disambiguate an ambiguous temporal expression; e.g., *if the meeting scheduled for next Wednesday has been moved forward two days, to which day has it moved?* This question is ambiguous, because the answer can be either Monday or Friday. Participants who had previously read ego-moving expressions tended to answer Friday (60% of responses), whereas those who had read time-moving expressions tended to answer Monday (69% of responses).

In a related study, Boroditsky and Ramscar (2002) examined whether thinking about movement through space influences how people interpret temporal expressions. They asked one group of participants to think of moving towards a certain point (which was in front of them), and asked another group of participants to think of moving a certain object towards them. Later, they presented these participants with the ambiguous question regarding Wednesday's meeting. Those participants who were earlier asked to think about moving forward tended to answer Friday (57% of responses), whereas participants who were asked to think of an object moving towards them tended to answer Monday (67% of responses).

7.3 Further considerations

The benefit of such methods is that they can demonstrate what constructs are cognitively accessible. However, the interpretation of such findings should be done with care. Specifically, an experimental finding demonstrating that thinking about one domain affects how people think about or behave towards another does not necessarily indicate that the two domains are based on common representational systems. For instance, Bargh and colleagues (Bargh & Chartrand 1999; Bargh, Chen, & Burrows 1996) found that participants that were shown films of elderly people later left the screening room more slowly than participants who had not seen such films. That is, thinking about the elderly caused people to walk more slowly. Clearly this does not mean that the dimensions of motion and age share a common representation. Instead, it shows that thinking about age may make certain constructs accessible (such as slowness), which can, in turn, affect behavior.

8. Summary

We have discussed some of the most common experimental methods used to study the interpretation of linguistic expressions in experimental settings. Some methods, such as lexical decisions, naming, and reading times rely on measuring behavior that occurs when participants are actively engaged with the experimental material. They are often referred to as "online measures" and are thought to index processes that are relatively low level and automatic rather than strategic ones. Other methods, such as those that rely on memory, fragment completion, or feature listing are "offline measures" as they rely on information (i.e., the dependent measure) that is collected well after the comprehension of the materials has been completed. These measures are therefore more susceptible to the working of strategic considerations employed by participants in the study.

Because our purpose was to introduce the reader to the different procedures, we did not elaborate on the more advanced technical considerations involved in each method. For those interested, Neely (1991) offers a comprehensive review of the different factors that affect lexical decision and naming tasks, Haberlandt (1994) reviews in detail different factors that affect performance on reading time and item-verification measures, and Keenan, Potts, Golding and Jennings (1990) provide a detailed review of different methodologies used to study inferences during reading.

References

Allbritton, D. W., McKoon, G., & Gerrig, R. J. (1995). Metaphor-based schemas and text representations: Making connections through conceptual metaphors. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 21*(3), 612–625.

Bargh, J. A. & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462–479.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality & Social Psychology, 71*(2), 230–244.

Becker, A. H. (1997). Emergent and common features influence metaphor interpretation. *Metaphor & Symbol, 12*(4), 243–259.

Black, J. B., Turner, T. J., & Bower, G. H. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning & Verbal Behavior, 18*(2), 187–198.

Blasko, D. G. & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*(2), 295–308.

Boroditsky, L. & Ramscar, M. (2002). The roles of body and mind in abstract thought. *Psychological Science, 13*(2), 185–189.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology, 3*(2), 193–209.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12–21.

Clark, H. H. & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.

Clark, H. H. & Clark, E. V. (1977). *Psychology and language. An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Costello, F. J. & Keane, M. T. (2001). Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*(1), 255–271.

Coulson, S. & Matlock, T. (2001). Metaphor and the space structuring model. *Metaphor & Symbol, 16*(3–4), 295–316.

Estes, Z. (2003). A tale of two similarities: Comparison and integration in conceptual combination. *Cognitive Science, 27*(6), 911–921.

Fauconnier, G. (1994). *Mental spaces: Aspects of meaning, construction in natural language*. New York, NY: Cambridge University Press.

Fauconnier, G. (1997). *Mappings in thought and language*. Cambridge, U.K. New York, NY: Cambridge University Press.

Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech, 9,* 217–227.

Gagne, C. L. (2001). Relation and lexical priming during the interpretation of noun-noun combinations. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*(1), 236–254.

Gentner, D. & Bowdle, B. F. (2001). Convention, form, and figurative language processing. *Metaphor & Symbol, 16*(3), 223–247.

Gentner, D., Bowdle, B. F., Wolff, P., & Boronat, C. (2001). Metaphor is like analogy. In D. Gentner & K. J. Holyoak (Eds.), *Analogical mind: Perspectives from cognitive science* (pp. 199–253). The MIT Press.

Gibbs, R. W. (2000). Making good psychology out of blending theory. *Cognitive Linguistics, 11*(3/4), 347–358.

Giora, R., Balaban, N., Fein, O., & Alkabets, I. (2005). Negation as positivity in disguise. In H. L. Colston & A. Katz (Eds.), *Figurative language comprehension: Social and cultural influences* (pp. 233–258). Hillsdale, NJ: Erlbaum.

Giora, R & Fein, O. (1999). Irony comprehension: The graded salience hypothesis. *Humor, 12*(4), 425–436.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language, 26*(1), 69–83.

Grady, J. (2000). Cognitive mechanisms of conceptual integration. *Cognitive Linguistics, 11*(3/4), 335–345.

Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1–31). San Diego, CA: Academic Press, Inc.

Hampton, J. A. (1987). Inheritance of attributes in natural concept conjunctions. *Memory & Cognition, 15*(1), 55–71.

Hasson, U. & Glucksberg, S. (2006). Does negation entail affirmation? The case of negated metaphors. *Journal of Pragmatics, 38,* 1015–1032.

Hasson, U., Simmons, J. P., & Todorov, A. (2005). Believe it or not: On the possibility of suspending belief. *Psychological Science, 16*, 566–571.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*(2), 141–154.

Horton, W. S. & Rapp, D. N. (2003). Out of sight, out of mind: Occlusion and the accessibility of information in narrative comprehension. *Psychonomic Bulletin & Review, 10*(1), 104–109.

Johnson, A. T. (1996). Comprehension of metaphors and similes: A reaction time study. *Metaphor & Symbolic Activity, 11*(2), 145–159.

Johnson, C. & Keil, F. (2000). Explanatory knowledge and conceptual combination. In F. C. Keil & R. A. Wilson (Eds.), *Explanation and cognition* (pp. 327–359). Cambridge, MA: The MIT Press.

Just, M. A. & Carpenter, P. A. (1976). The relation between comprehending and remembering some complex sentences. *Memory & Cognition, 4*(3), 318–322.

Kaup, B. (2001). Negation and its impact on the accessibility of text information. *Memory & Cognition, 29*(7), 960–967.

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading?: A question of methodologies. In D. Balota, G. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading*. Hillsdale, NJ: Erlbaum.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2), 259–284.

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Stanford, CA: Stanford University Press.

MacDonald, M. C. & Just, M. A. (1989). Changes in activation levels with negation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15*(4), 633–642.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am guilty" vs. "I am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology, 40*(4), 443–449.

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition, 78*(1), B17–B25.

McGlone, M. S. & Harding, J. L. (1998). Back (or forward?) to the future: The role of perscpective in temporal language comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*(5), 1211–1223.

McKoon, G. & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 440–466.

McKoon, G. & Ratcliff, R. (1992). Inference during reading. *Psychological Review, 99*(3), 440–466.

McKoon, G. & Ratcliff, R. (1995). Conceptual combinations and relational contexts in free association and in priming in lexical decision and naming. *Psychonomic Bulletin & Review, 2*(4), 527–533.

Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90*, 227–234.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Nisbett, R. E. & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Peterson, L. & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.

Richards, A. & French, C. C. (1991). Effects of encoding and anxiety on implicit and explicit memory performance. *Personality & Individual Differences, 12*(2), 131–139.

Searle, J. (1979). Metaphor. In A. Ortony (Ed.), *Metaphor & Thought* (pp. 92–123). New York, NY: Cambridge University Press.

Sloutsky, V. & Goldvarg, Y. (2004). Mental representation of logical connectives. *The Quarterly Journal of Experimental Psychology A, 57*(4), 636–665.

Smith, C. D. (1981). Recognition memory for sentences as a function of concreteness/abstractness and affirmation/negation. *British Journal of Psychology, 72*(1), 125–129.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition, 16*(4), 283–298.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory & Language, 47*(4), 530–547.

Whitney, P., Waring, D. A., & Zingmark, B. (1992). Task effects on the spontaneous activation of trait concepts. *Social Cognition, 10*(4), 377–396.

Wilkenfeld, M. J. & Ward, T. B. (2001). Similarity and emergence in conceptual combination. *Journal of Memory & Language, 45*(1), 21–38.

Eye movements in language and cognition

A brief introduction

Daniel C. Richardson, Rick Dale and Michael J. Spivey

1. The role of eye movements in the visual system

The eye independently evolved over 40 times in nature (Fernald 1997), yet, strikingly all animals with developed visual systems actively control their gaze using eye or head movements (Land 1995; Treue 2001). Indeed, it can be argued that the most frequent behaviour of human beings is movement of the eyes (Bridgeman 1992). This 'ceaseless twitching', as one early researcher described it (Stratton 1906), is the visual system's solution to the huge amount of available visual information and limited processing resources. The human eye covers a visual field of about 200° but receives detailed information from only 2° (Levi, Klein, & Aitsebaomo 1985). This tiny high-resolution area of the retina, which receives input from an area of the visual field about the size of a thumbnail at arm's length, is called the fovea, and is jerked around at speeds of up to 500° a second, during which its sensitivity drops to near blindness levels (Matin 1974; Thiele, Henning, Kubischik, & Hoffmann 2002). During the 200–300 milliseconds it is at rest, however, over 30,000 densely packed photoreceptors in the fovea provide high acuity color vision. Eye movements direct this information conduit to relevant portions of the world, and are therefore fundamental to the operation of the visual system.

There are several classes of eye movements. Of most interest to the cognitive psychologist – and arguably most common – are saccades, the rapid, ballistic movements that move the eye around the visual field roughly 3–4 times a second. Other classes of eye movements, such as smooth pursuit, vergence, optokinetic nystagmus, torsion, and micro-saccades, serve to maintain fixation despite head, body or object motion, changes in depth, and to correct for muscle drift and inaccuracy. They are explored for purposes of understanding the oculomotor system, in and of itself, but rarely as a measure of cognitive processing. Typically, experimental psychologists measure when saccades are launched, where they land, and how long the eye stays there. The pattern of locations that saccades visit is termed the 'scanpath', the duration between a stimulus onset and a saccade is called the 'saccade latency', and the amount of time spent looking at a particular location is called the 'fixation duration'. In this chapter we will show how eye movement metrics such as these have yielded a rich skein of data relating to cognitive processes, and suggest ways that they can be further mined by experimental psychologists.

2. The value of eye movements to cognitive psychologists

Eye movements are uniquely poised between perception and cognition. They are central to the function of the visual system, but for such scanning to be efficient, it cannot be simply a random sample of the visual world. To be useful, eye movements must be related to an organism's memories, expectations and goals. Consequently, eye movements are driven equally by bottom-up perceptual properties of the world and top-down cognitive processes. This role in the perception-action cycle makes saccadic behaviour particularly informative for the experimental psychologist.

There are several specific characteristics of eye movements which also prove to be of great practical and theoretical benefit. Saccades occur roughly 3–4 times per second. During the response time of a typical experimental task then, eye-tracking data can provide a semi-continuous record of regions of the visual field that are briefly considered relevant for carrying out an experimental task. Crucially, this record provides data during the course of cognitive processing, not merely after processing is complete, as is often the case with more conventional measures. Eye-tracking data thus provide not only behavioral end products of our cognitive processes but also clues to the process through which they are achieved. Importantly, this sensitive semi-continuous measure of cognitive processing can also be used in ways that do not interrupt task processing with requests for metacognitive reports or other overt responses. Thus eye-tracking allows for a certain degree of ecological validity in task performance, as the responses it collects are ones that typically occur regardless of experimenters' instructions and participants' intent.

Moreover, eye movements exhibit a unique sensitivity to partially active representations that may not be detectable by most other experimental measures, or even result in any other overt behaviour. Since eye movements are extremely fast, quickly corrected, and metabolically cheap, compared to other motor movements, they have a much lower threshold for being triggered. Hence, briefly partially-active representations – that might never elicit reaching, speaking, or even internal monolog activity because they fade before reaching those thresholds – can nonetheless occasionally trigger an eye movement that betrays this otherwise-latent momentary consideration of that region of the visual display as being potentially relevant for interpretation and/or action. For example, in a classic experiment in psychology, two speech sounds that vary continuously between "ba" and "pa" are categorically perceived by participants, who report hearing *either* a "ba" or a "pa" and respond by pressing a corresponding button (Liberman, Harris, Hoffman, & Griffith 1957). However, McMurray and colleagues (McMurray, Tannenhaus, Aslin, & Spivey 2003) showed that eye movements between the two response buttons increased when the speech sound was near the ba/pa boundary. Thus, the graded nature of the perceptual process, which is lost in a categorical response, is revealed in the time course of eye movements.

Eye movements have a long and successful history as a window into perceptual and cognitive processing. The following sections present a subset of that research that would be of particular interest to cognitive linguists (for a broader review, see Richardson & Spivey 2004a, 2004b). First, we will briefly describe how eye movements reveal psychological processes in everyday tasks of perception and memory. Then we will show how very similar cognitive processes produce similar eye-movement patterns in more 'offline' situations, where relevant visual stimuli may not even be present. We shall then turn to the many ways in which eye movements reveal facts about language processing, from the mechanics of reading, to the integration of visual and verbal information, to the conceptual representation of narratives and metaphors. This chapter will conclude with a discussion of some of the practical methodological necessities involved in designing, conducting, and analyzing eyetracking experiments.

3. Perception and action

Visual attention is not always coincident with eye position. Posner, Snyder, and Davidson (Posner 1980) demonstrated that participants' covert visual attention can be dissociated from the fovea when they are explicitly instructed to not move their eyes. However, it is highly likely that spatial attention and saccade planning are closely coupled during natural unconstrained eye movement (Findlay & Gilchrist 1998). There is behavioural evidence that covert attention directed in one direction can lead to deviations in orthogonal saccades (Sheliga, Riggio, Craighero, & Rizzolatti 1995), and neuropsychological evidence from single cell recordings suggesting that they utilize overlapping neural systems (Corbetta, Akbudak, Conturo, Snyder, Ollinger, & Drury 1998). Moreover, planning a saccade toward a location improves processing at that location, regardless of whether or not the saccade is launched (Hoffman & Subramaniam 1995; Sheliga, Riggio, & Rizzolatti 1994; Shepherd, Findlay, & Hockey 1986), and indeed, evidence shows that microstimulation of neurons in the frontal eye fields can cause both a saccade to a certain location (Robinson & Fuchs 1969), and, with a lower level of stimulation, an absence of eye movement, but improved stimulus detection at that location (Moore & Armstrong 2003; Moore & Fallah 2001).

How are eye movements and visual attention directed around a visual stimulus? In general, when viewing a static scene, the eyes appear to be driven by both visual properties of the stimulus and top-down effects of knowledge and expectations (Henderson 2003). For example, Buswell (1935) showed that a viewer will pay scant attention to solid regions of colour in a painting, and instead will tend to fixate regions of contrast and high spatial frequency; top-down effects will be seen in the viewers' saccades to semantically important regions of the painting, such as faces, and the ways in which a naïve viewer will inspect the painting differently from an art expert. These findings have been recently replicated on a large scale. Wooding and colleagues installed an autonomous eye tracker in a public museum in London, and collected data from over 5000 subjects looking at works from the National Gallery (Wooding 2002; Wooding, Muggelstone, Purdy, & Gale 2002). They too found that only a small set of regions in a work of art were reliably fixated by viewers.

We in live a world more dynamic and interactive than an art gallery, however. How are eye movements integrated with action in the course of everyday tasks? With the advent of headband-mounted eye-tracking, which allows natural movement of the entire body, this real-time measure of perceptual and cognitive processing has been applied to a number of more richly interactive and ecologically valid experimental tasks and paradigms. Eye movements can even reveal the everyday strategies we employ while carrying out basic tasks, such as making a sandwich. For example, Land and Hayhoe (2001) found that eye movements are tightly linked with moment-to-moment goals and sub-tasks. Task-related fixations illuminating visual memory processes have been examined in detail using the block-copying task developed by Ballard, Hayhoe and colleagues (Ballard, Hayhoe, & Pelz 1995; Ballard, Hayhoe, Pook, & Rao 1997; Hayhoe, Bensinger, & Ballard 1998).

Ballard et al. (1995) recorded participants' eye movements during a block-pattern copying task. Participants had a model pattern, a resource of blocks, and a workspace in which to copy the model. The participants' hand actions were recorded, and a headband-mounted eye tracker recorded their eye movements to obtain a window on the strategy used in the task. One method participants could use is to look at the model area and memorize the pattern; each block could then be located in the resource area and placed in the workspace. A second, less memory-intensive, method would be to remember the color and location of one block from the model, collect it from the resource, place it in the workspace, and then consult the model again for the next block. The strategy used by participants, however, most often entailed the minimal possible memory demands. Participants would commonly fixate the model, then fixate and pickup a correctly colored block from the resource area, fixate the model yet again, and then place the block in the workspace. Thus, two fixations per block were made on the model – one to extract color information, one to extract relative spatial location information.

Eye movements reveal a cognitive process of "indexing," whereby the location of an object is maintained in working memory, and other properties can be "looked up" in the environment as they are needed, moment by moment, during a task (Ballard et al. 1997). For example, in a computerized, gaze-contingent version of the block-copying task, the color of a block was changed during a saccade (Hayhoe et al. 1998). The participants rarely noticed this property change, demonstrating that they had not encoded the information, but instead relied upon the fact that an eye movement could access it when required.

4. Cognition

A general case can be made that 'offline' cognitive processes such as remembering, imagining and reasoning may employ many of the same mechanisms as 'online' perceiving and acting in the world (Barsalou 1999; Damasio 1989; Kosslyn, Behrmann, & Jeannerod 1995; Martin 2001; Ryle 1949). Certainly, we will see here that eye movement patterns during cognitive activity bear a striking resemblance to those during the perception and manipulation of objects in the world. This continuity between perception and cognition can be exploited by psychologists, who can use overt eye-movement behaviour to investigate internal mental processes.

A clear example of this parallel exists between Ballard et al's (1995) task, where subjects moved blocks around, and a set of experiments where subjects remembered a series of verbally presented facts. Richardson and Spivey (2000) presented four talking heads in sequence, in the four quadrants of the screen, each reciting an arbitrary fact and then disappearing (e.g., "Shakespeare's first plays were historical dramas. His last play was The Tempest."). With the display completely blank except for the lines delineating the four empty quadrants, a voice from the computer delivered a statement concerning one of the four recited facts, and participants were instructed to verify the statement as true or false (e.g., "Shakespeare's first play was The Tempest.").

While formulating their answer, participants were twice as likely to fixate the quadrant that previously contained the talking head that had recited the relevant fact than any other quadrant. Despite the fact that the queried information was delivered auditorily, and therefore cannot possibly be visually accessed via a fixation, participants systematically fixated blank regions of space. This result was replicated when the talking heads were replaced by four identical spinning crosses.

Moreover, in a "'tracking' condition" participants viewed the grid through a virtual window in the center of the screen. Behind this mask, the grid moved, bringing a quadrant to the center of the screen for fact presentation. Then, during the question phase, the mask was removed. Even in this case, when the spinning crosses had all been viewed in the center of the computer screen, requiring no eye movements, and the relative locations of the quadrants implied by translation, participants continued to treat the quadrant associated with the queried fact as conspicuously worthy of overt attention. In fact, even if the crosses appear in empty squares which move around the screen following fact delivery, participants spontaneously fixate the square that was associated with the fact in its new location (Richardson & Kirkham 2004). The behaviour of associating events and information with a moving location, and re-fixating that location when the information is relevant has been termed 'spatial indexing'. Remarkably, there is evidence of spatial indexing behaviour in the eye movements of infants as young as 6 months of age (Richardson & Kirkham 2004).

When subjects listened to pieces of semantic information, they associated them with spatial indexes, just as the participants in Ballard and colleagues' block moving task did for the blocks they were manipulating (Richardson & Spivey 2000; Spivey, Richardson, & Fitneva 2004). As many researchers have argued (Ballard et al. 1997; O'Regan 1992; Pylyshyn 1989, 2001), deictic pointers can be used in visuomotor routines to conserve the use of working memory. In Brooks' (1991) words, the 'world can be used as its own best representation'. Instead of storing all the detailed properties of an object internally, one can simply store an address, or pointer, for the object's location in the environment, via a pattern of activation on an attentional/oculomotor salience map in parietal cortex (Duhamel, Colby, & Goldberg 1992), along with a spatial memory salience map in prefrontal cortex (Chafee & Goldman-Rakic 1998, 2000; Goldman-Rakic, Chafee, & Friedman 1993). If this spatial pointer is associated with some kind of coarse semantic information, e.g., a pattern of activation in one of the language cortices, or auditory cortex, or even visual cortex, then the spatial pointer can be triggered when sensory input activates that semantic information. Such pointers allow the organism to perceptually access relevant properties of the external world when they are needed.

It actually should not be surprising that an embodied working memory system using deictic pointers would attempt to index information from events that are over and done with. The pointer doesn't "know" that the sought-after information at its address is long gone precisely because it has offloaded that knowledge onto the environment – it wouldn't be a pointer otherwise. These eye movement findings demonstrate the robustness and automaticity with which spatial indices are relied upon in order to employ the body's environment as sort of noticeboard of 'virtual post-it notes' that complement our internal memory.

However, many complex tasks we face on a daily basis do not necessarily involve indexing of relevant objects in a task space. For example, producing/hearing descriptions of far away scenes or events, gossip about people who are absent, and discussions of abstract concepts, do not involve explicit reference to visible elements of the immediate situational context. An important question concerns the extent to which eye movements may be indicative of imagery processes when carrying out such tasks. How are eye movements implicated in visualizing a complex story or description? Will scanning of the visuo-spatial backdrop that is available to a listener be at all relevant during comprehension of language that refers to things that are not visually co-present with the speech?

In a headband-mounted eye-tracking experiment, Spivey and Geng (experiment 1, 2001; see also Spivey, Tyler, Richardson, & Young 2000) recorded participants' eye movements while they listened to spoken descriptions of spatiotemporally dynamic scenes and faced a large white projection screen that took up most of their visual field. For example, "Imagine that you are standing across the street from a 40-story apartment building. At the bottom there is a doorman in blue. *On the 10th floor, a woman is hanging her laundry out the window. On the 29th floor, two kids are sitting on the fire escape smoking cigarettes. On the very top floor, two people are screaming*." While listening to the italicized portion of this passage, participants made reliably more upward saccades than in any other direction. Corresponding biases in spontaneous saccade directions were also observed for a downward story, as well as for leftward and rightward stories. (A control story, describing a view through a telescope that zooms in closer and closer to a static scene, elicited about equal proportions of saccades in all directions). Thus, while looking at ostensibly nothing, listeners' eyes were doing something similar to what they would have done if the scene being described were actually right there in front of them. Instead of relying solely on an internal "visuospatial sketchpad" (Baddeley 1986) on which to illustrate their mental model of the scene being described, participants also recruited the external environment as an additional canvas on which to depict the spatial layout of the imagined scene.

Although eye movements may not be required for vivid imagery (Hale & Simpson 1971; Ruggieri 1999), it does appear that they often accompany it (Antrobus & Antrobus 1969; Brandt & Stark 1997; Demarais & Cohen 1998; Hebb 1968; Laeng & Teodorescu 2002; Neisser 1967). Early empirical investigations found that the frequency of eye movements increases during mental imagery, particularly that of a spatial nature (Clark 1916; Goldthwait 1933; Perky 1910; Stoy 1930; Totten 1935); and an increase in rapid fluttering of the eyes while sleeping correlates with vividness of dreams (Antrobus & Antrobus 1969; Goodenough, Shapiro, Holden, & Steinschriber 1959; Roffwarg, Dement, Muzio, & Fisher 1962). But what is it that the eyes are trying to do in these circumstances? Obviously, it is not the case that the eyes themselves can actually externally record this internal information. When the eyes move upward from the imagined 10th floor of the apartment building to the imagined 29th floor, no physical mark is left behind on the external location in the environment that was proxying for that 10th floor.

In the case of Spivey and Geng's (2001) eye movements during imagistic spoken narrative comprehension, a few pointers allocated on a blank projection screen will obviously not make reference to any external visual properties, but they can still provide perceptual-motor information about the relative spatial locations of the internal content associated with the pointers (see also Altmann & Kamide 2004). If one is initially thinking about *x* (e.g., the 10th floor) and then transitions to thinking about *y* (e.g., the 29th floor), then storing in working memory the relation *above* (*y*,*x*) may not be necessary if the eye movements, and their allocation of spatial indices, have embodied and externalized that spatial relationship in the environment already (cf. Pylyshyn 1989). In this way, a "low-level" motor process, such as eye movements, can actually do some of the work involved in the "high-level" cognitive act of representing spatial relations in visual imagery elicited by linguistic input. Eye movement data thus reveal a powerful demonstration of how language about things not co-present is interfaced with perceptual-motor systems that treat the linguistic referents as if they were co-present.

It seems clear from the evidence presented here that eye movements are a rich source of information about cognitive processing, even when the relevant items are not physically present, but are recalled from memory or merely imagined. Although some researchers today argue, on the basis of null results, that eye movements are not really indicative of cognitive processes at all (Anderson, Bothell, & Douglass 2004), other researchers are demonstrating that eye movements can reveal not just which cognitive representation might be active, but how they are being manipulated. For example, eye movements appear to have a relationship to the reasoning process in mechanical problem solving (Hegarty 1992; Hegarty & Just 1993; Hodgson, Bajwa, Owen, & Kennard 2000; Rozenblit, Spivey, & Wojslawowicz 2002) and insight problem solving (Grant & Spivey 2003; Jones 2003; Knoblich, Ohlsson, & Raney 2001).

5. Language

Language processing encompasses a spectrum of phenomena, from the largely perceptual aspects of word identification, to the largely conceptual aspects of metaphor understanding. As one might imagine from the preceding sections, eye movements can provide insight at each of these levels. The most apparent, and most studied, link between eye movements and language is in the process of reading.

The general characteristics of eye movements during reading have been studied in great depth over the past quarter century (for thorough reviews, see Rayner 1978, 1998). This methodology has revealed a number of important facts about how people's eyes move when they read. For example, the eyes rest in fixation for approximately 200–250 milliseconds during reading. Saccades between fixations span an average about 2 degrees of visual angle, although this is better expressed here in terms of a span of 7 to 9 letter spaces, since the number of letters covered remains largely invariant despite differences in text size or distance (Morrison & Rayner 1981). The chances of an individual word being fixated vary according to whether it is a content word (85%) or a function word (35%) (Carpenter & Just 1983), and in relationship to the length of the word, with 2–3 letter words being skipped 75% or the time, but 8 letter words fixated almost always (Rayner & McConkie 1976). Eye movements also vary as a function of the syntactic and conceptual difficulty of the text (Ferreira & Clifton 1986; Rayner, Sereno, Morris, Schmauder, & et al. 1989). Although readers typically move their eyes forward when reading, approximately 10–15% of saccades move backward, fixating previous letters or words. These regressive saccades are thought to be related to difficulties in processing an individual word, or difficulties in processing the meaning or structure of a sentence; in these cases, readers can often accurately re-fixate the part of the text that generated confusion (Murray & Kennedy 1988).

These features of eye movements during reading – gaze durations, saccade lengths, occurrence of regressions, and a number of variations on these measures – can be used to infer moment-by-moment cognitive processing of a text by the reader (Just & Carpenter 1980; Rayner et al. 1989). Details of the cognitive processes of pronoun resolution and co-reference, word frequency, lexical ambiguity, syntactic ambiguity, as well as the influence of semantic and discourse context on these processes, can all by gleaned from analyses of eye-movement patterns (Rayner 1998; Tanenhaus & Trueswell 1995).

Light, headband mounted eye trackers have allowed researchers to extend the online measurement of language processing beyond reading, to the perception and understanding of spoken language in a rich, naturalistic visual context. One field of research begins with the feature of eye movements, noted above, that participants will often look briefly at an object that is initially considered relevant for action, and then quickly re-fixate their eyes on another object that becomes the actual target of the action. This feature has been exploited to study many factors in the time course of speech processing and language understanding.

For example, Spivey and colleagues (Spivey-Knowlton, Tanenhaus, Eberhard, & Sedivy 1998) sat participants in front of a display of objects such as a candle, bag of candy, a pencil, and a spoon. The participants were then instructed to "Pick up the candy,". About a third of the time participants fixated the candle for a couple hundred milliseconds before looking to and reaching for the candy. Participants typically denied looking to the candle at all, and yet their eye movements revealed a process substantially different from their conscious report and their manual action. This kind of brief interference between similar sounding object names occurs not just for cohorts but also for rhymes (Allopenna, Magnuson, & Tanenhaus 1998), as well as for novel words from an artificial lexicon (Magnuson, Tanenhaus, Aslin, & Dahan 2003), and even for words that sound similar across two different languages (Marian & Spivey 2003; Spivey & Marian 1999). It appears that the acoustic uptake of spoken input is continuously mapped onto visually relevant lexical representations, such that partial phonological matches to the names of multiple visual objects induces competition between partially active representations, in a system something like interactive processing in the TRACE connectionist model of spoken word recognition, (Elman & McClelland 1988; Magnuson, McMurray, Tanenhaus, & Aslin 2003; McClelland & Elman 1986).

A similar influence of visual context is observed with temporary ambiguities that arise across words, in the syntax of a sentence. When presented with a display containing an apple on a towel, another towel, and an empty box, and then instructed to "Put the apple on the towel in the box," participants often looked briefly at the irrelevant lone towel near the end of the spoken instruction before returning their gaze to the apple, grasping it, and then placing it inside the box (Spivey, Tanenhaus, Eberhard, & Sedivy 2002; Tanenhaus, Spivey Knowlton, Eberhard, & Sedivy 1995). (With unambiguous control sentences, such as "Put the apple that's on the towel in the box," they almost never looked at the irrelevant lone towel.) In this case, the syntax is ambiguous as to whether the prepositional phrase "on the towel" is attached to the verb "put" (as a movement destination) or to the noun "apple" (as a modifier). Given the actions afforded by the display, the latter syntactic structure is the correct one. However, people tend to have a bias toward interpreting an ambiguous prepositional phrase as attached to the verb (Rayner, Carlson, & Frazier 1983), at least when it is an action verb like "put" (Spivey-Knowlton & Sedivy 1995). Thus, the brief fixation of the irrelevant lone towel indicates a temporary partially-activated incorrect parse of the sentence. To demonstrate the influence of visual context on this syntactic ambiguity resolution process, the display was slightly altered to include a second apple (resting on a napkin). In this case, the visual co-presence of the two potential referents for the phrase "the apple" should encourage the listener to interpret the ambiguous prepositional phrase "on the towel" as a modifier (in order to determine which apple is being referred to) rather than as a movement destination (Altmann & Steedman 1988; Spivey & Tanenhaus 1998). And, indeed, with this display, participants rarely fixated the irrelevant lone towel, indicating that visual context had exerted an immediate influence on the incremental syntactic parsing of the spoken sentence (Knoeferle, Crocker, Scheepers, & Pickering 2005; Spivey et al. 2002; Tanenhaus et al. 1995).

The word-by-word interfacing between spoken language and visual perception is also evidenced by reference resolution with complex noun phrases. Eberhard, Spivey-Knowlton, Sedivy, and Tanenhaus (1995) presented participants with a display of blocks of various shapes, colors, and markings, and gave them instructions like "Touch the starred yellow square." When the display contained only one starred block, participants often fixated on the target block before the head noun of the noun phrase had even been spoken. Fixation of the target block was slightly later when the display contained another starred block that was not yellow, and later still when the display also contained a starred yellow block that was not a square. This result shows that even before hearing the noun that refers to the object being described, listeners are processing the pre-nominal adjectives as they are heard and mapping their meaning onto the options available in the visual context.

More recently, researchers have employed eye movement techniques to show that listeners are remarkably sensitive to subtle aspects of language, and employ that information in directing their gaze around a display. For example, Altmann and Kamide (2004) have demonstrated that participants will fixate a cake before hearing the word spoken in the sentence, 'The boy will eat the cake'. This anticipatory saccade will not occur when subjects hear, 'The boy will move the cake'. This evidence demonstrates participants are activating rich 'thematic role' knowledge (Ferretti, McRae, & Hatherell 2001) of the verb "eat", and fixating likely candidates for this action before the word is spoken.

Matlock and Richardson (2004) have provided further evidence of the nuances of language that drive eye movements around a scene. In previous reading time studies, Matlock (2004) found evidence that readers would mentally simulate motion when reading sentences such as 'the fence runs along the garden'. Such use of figurative language is termed 'fictive motion', since although the motion verb 'run' is used, no literal motion takes place. Matlock and Richardson (2004) showed that listeners would look longer along the relevant path when they heard 'the fence runs along the garden', compared to 'the fence is around the garden'. This suggests that fictive motion, far from being an example of a 'dead metaphor', elicits something like a perceptual simulation (e.g., Barsalou 1999) and influences how a listener directs their attention across a picture.

Eye movements thus reveal the incremental and interactive nature of spoken language comprehension. Subjects are gradually influenced by the incremental delivery of linguistic information, and eye movements exhibit the continuous, partially active representations that arise during processing. In addition, eye movements have permitted the observation of powerful interactive effects between language and vision. It seems that this incremental process of language comprehension can be strongly constrained by appropriate visual contexts, and that moment-by-moment visual perception can be driven by subtle aspects of language such as thematic roles and figurative motion.

6. Eye movement methodology

We hope that the reader is convinced by now that recording eye-movements does indeed provide a unique source of data for constraining one's theories about language and cognition. The ways in which the visual environment constrains, and is used by, various linguistic and cognitive processes are becoming better understood due to the insights afforded by many findings from eye-tracking experiments. However, there are a number of safeguards and practical tips that one accumulates over years of experience with eyetracking studies that are worth considering before a newcomer dives right into collecting a large mass of eye-movement data (for reviews, see Rayner 1998; Tanenhaus & Spivey-Knowlton 1996).

This section of the chapter is aimed at providing some concrete methodological preparation for students and researchers interested in tracking people's eye movements as a measure of cognitive processes. Of course, it would be naïve of us to think that a chapter, by itself, could prepare a reader to successfully and accurately track the eye movements of experimental participants. The only way to get good at eyetracking is to receive hands-on training from someone with years of experience, and then practice, practice, practice. That said, perhaps some of the following advice could speed up that learning process.

6.1 Choice of eyetracker

The first decision to make in considering eye movements as an experimental measure is what aspect of oculomotor behavior one thinks may be especially informative for the aspect of cognition being tapped. It is important to imagine in advance exactly what eye-movement analysis will be conducted and what the ideal results (and statistical analysis) will look like. This will often determine what kind of eyetracker is most appropriate for your experiment. Luckily for your experimental participants, the facility with which eye movements are tracked has improved considerably over many decades of developing technology. Devices attached directly to the eyeball (Delabarre 1898; Huey 1898; Yarbus 1965) gave way to photographic techniques (Diefendorf & Dodge 1908; Tinker 1928), which in the last few decades, have been replaced by electronic detection of the small differences among reflective properties of the eye (Cornsweet & Crane 1973; Merchant, Morrissette, & Porterfield 1974; Young 1970) that permit rapid and accurate calculation of gaze direction (for a review of the history of eye tracking methods, see Richardson & Spivey 2004a). There are also search coils (typically surgically implanted in monkeys) and electro-oculogram (EOG, for measuring movement latency, but not eye position), but those will not be discussed here as they are less practical for cognitive tasks with humans.

For the last few decades, uses of eyetracking in reading have tended to focus on durations of fixations on various words (and numerous permutations of this metric). When a sentence becomes syntactically or semantically complex, ambiguous, or misleading, fixations on the words that cause or resolve that confusion will often be longer than fixations on other words (of equal length and lexical frequency). The frequency of regressive saccades, back to earlier portions of a sentence, are also used as a measure of processing difficulty. Similar kinds of measures have been used for measuring cognitive processes during scene viewing. Two kinds of table-mounted eyetrackers have typically been used for these purposes. A dual-Purkinje eyetracker – rather expensive, but with extremely high spatial and temporal resolution – points infrared light into the eyeball and records the reflections off the front surface of the cornea (first Purkinje image) and the back surface of the lens (fourth Purkinje image). These two reflections allow for a calculation of the point-of-regard of the eye, mapping what part of the computer screen (on which the sentences are presented) is being foveated. Limbus tracking uses infrared emitter diodes and infrared detector diodes to record the position of the boundary between the iris and the sclera (white of the eye). Figure 1 shows an image of the Dr. Bouis limbus tracker. Limbus trackers are less expensive than dual-Purkinje eyetrackers, but also tend to have lower spatial and temporal resolution, as well as performing poorly along the vertical axis (since the eyelids can obscure the upper and lower portions of the iris-sclera boundary). Therefore, using limbus tracking for measuring reading often requires presenting only one sentence at a time, and is rarely used for complex scene viewing.

While they both tend to have higher resolution than other eyetracking systems (fractions of a degree of visual angle, allowing one to know *which letter* is being fixated), two important limitations of dual-Purkinje and limbus trackers are that they are generally only used with computerized displays and with the head held motionless. In some cases, a chin and forehead rest are sufficient, but it is often necessary to also use a bitebar. A bitebar is

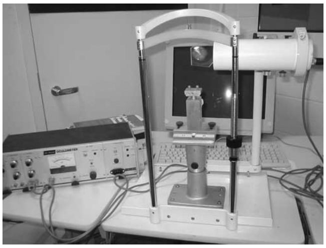


Figure 1. A Dr. Bouis limbus eyetracker, set up for recording movements of the right eye during reading. The metal pedestal in the middle is where the bitebar would be bolted in place. The glass eye piece above the pedestal is a silvered mirror that reflects the infrared light from the emitter diodes to the eye and back to the detector diodes, but allows normal visible light to pass, so that the participant can see through it.

Figure 2. An ISCAN, Inc. remote video-based eyetracker, set up for recording movements of one or the other eye during viewing of simple scenes on a computer screen. The cylinder in between the monitor and the camera lens is the infrared emitter. The mechanism immediately below the camera itself allows it to tilt up and down and pan left and right to accommodate small head movements.

made by heating dental wax in hot water until it is soft and malleable, shaping the wax around a horseshoe-shaped metal plate, re-warming the wax, and having the experimental participant bite down on the soft wax so that it makes a dental impression on top and underneath. (The third author once had a participant express reticence about this procedure, but when she was invited to quit the experiment, she said, "No, I'll do it, but can you close the lab door? I don't want any of my friends to walk by and see me doing this.") When the wax cools again and hardens, the plate is bolted onto a post mounted on the table under the eyetracker's infrared-emitting apparatus. While the participant is looking at the computer screen, and reading sentences or inspecting complex scenes that are typically extracted from any situational context, with the room lights off, they are biting down on the form-fitted bitebar so that their head does not move more than a millimeter or two. And the experimenter hopes to get data that are somehow ecologically valid.

Remote video-based eyetrackers (Figure 2) have an infrared emitter that bathes the participant's face in low-intensity infrared light, and the eye camera zooms in to one of the eyes and sends the infrared image to a computer. The eye camera is on a swivel that allows it to re-orient its aim if the participant's head moves slightly, either automatically or by remote control. The computer's software performs some image processing on the

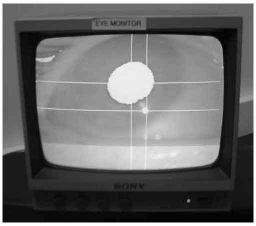


Figure 3. Image of the eye from a video-based eye camera, after the computer's image processing has placed crosshairs on the center of the pupil (upper left) and on the corneal reflection (lower right). (When these crosshairs are stable, and not jittery, the estimate of point-of-regard can be reliable and accurate.)

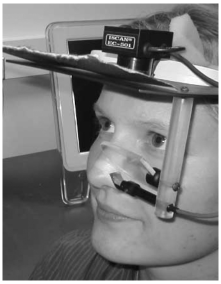


Figure 4. An ISCAN, Inc. headband-mounted eyetracker, set up for recording movement of the left eye during natural interactive task performance. Under the brim of the headband, an emitter sends infrared light down to the slivered mirror, which reflects it into the eye. Then the light bounces back and up to the eye camera, adjacent to the emitter, providing an image of the eye like that in Figure 3. The black cylinder underneath the silvered mirror is the scene camera whose image of the participant's field of view is reflected off the outside portion of the silvered mirror, off another mirror, and then into its lens.

eye camera's signal, finding the large disc of low reflectance (for "dark pupil trackers") or high reflectance (detecting light reflected back from the eye's tapetum, for "bright pupil trackers"). This disc is assumed to be the pupil, and crosshairs are placed at its center, as seen in Figure 3. The infrared emitter's reflection off the cornea is also identified in the image of the eye (Figure 3). With these two points on the eye's surface identified, and a generic model of the curvature of the eye, the software can estimate the direction of the eye relative to the head. With relatively minor head movement, one can calibrate these eye positions onto computer screen positions, interpolate between calibrated points, and track where on the screen the participant is looking, with an accuracy better than one degree of visual angle.

Eyetrackers that are yoked with computer-screen stimulus presentation can often allow for automated data analysis (scan paths, fixation durations, etc.) either by adding an electromagnetic head-tracker that subtracts out head-movement in determining where in the field of view the eyes are pointed, or by using a head-mounted scene camera (Figure 4) and placing bright markers on the corners of the computer screen so that the calibration routine can re-normalize the eye-position estimate to wherever the computer screen's reference frame is in the scene camera's view. However, when the eyetracker is not yoked with computer-screen stimulus presentation, but instead eye movements are to be recorded during natural visuomotor tasks involving several 3-D objects, automated data analysis will often be infeasible because the locations of objects on the scene camera will move substantially during head and trunk movements. The eye-position estimate will still be accurate in terms of coordinates in the scene camera's reference frame, but having an automated algorithm that knows where the different objects are in that moving reference frame can prove difficult. Headband-mounted eyetrackers that allow this kind of natural head and body movement (Figure 4) often require hand-coded frame-by-frame data analysis of videotaped recordings of eye-position superimposed (as crosshairs) on the scene camera's view. This will usually require a digital VCR.

As their precision and accuracy are somewhat lower than the Purkinje-image eyetrackers or even the limbus eyetrackers, these video-based remote and head-mounted eyetrackers are less frequently used for studying reading, where *seconds of arc*, rather than degrees or minutes, are crucial. These kinds of eyetrackers (manufactured by a variety of companies, including ISCAN, Inc., SR Research, and Applied Science Laboratories) are more typically used for experiments where the visual display will be carved up into no more than a dozen or so regions of interest. For example, one could measure what critical objects in a participant's environment are fixated, for how long and in what order, while he/she is driving on a curvy road (Land & Lee 1994), following spoken instructions to move objects around (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus 1995), making critical decisions in a chess game (Charness, Reingold, Pomplun, & Stampe 2001), selecting response buttons during a categorical speech perception task (McMurray, Tanenhaus, Aslin, & Spivey 2003), verbally describing a visual scene (Griffin & Bock 2000), attempting to solve a diagrammatic problem (Grant & Spivey 2003), following a pre-recorded baking recipe (Chambers, Tanenhaus, & Magnuson 2004), or even just preparing a peanut butter and jelly sandwich (Land & Hayhoe 2001).

6.2 Calibration

The first thing to practice for actually conducting an eyetracking session is calibration. You need to make sure the eye camera (or the infrared diodes, for a limbus tracker) is evenly lined up on the eye such that when the participant looks at the edges of the main viewing field, the pupil (or iris, for a limbus tracker) is still in full view of the tracker. During calibration, several eye-position signals will be mapped onto particular regions of the participant's field of view (often the four corners, plus some others). To maintain a linear relationship between these positions, the participant's head must be as immobile as possible during the minute-or-two of the actual calibration routine. The faster you can go through asking them to fixate each calibration point on the screen, and collecting the eye-position signal belonging to that location (watch out for blinks!), the less time there will be for the head to drift, and the more linear your calibration points' relationships to one another will be. With a good set of evenly spaced calibration points, and a clean, stable eye-position signal associated with each of them, the software's interpolation for regions in between those calibration points should be reasonably accurate.

6.3 Display parameters

Perhaps the most important thing to keep in mind when designing an eye-tracking experiment with spoken linguistic input is to plan in advance exactly how the eye-position record will be coded. It is sometimes helpful to have an initial centralized fixation point where the subject is instructed to look at the beginning of each trial; and critical relevant objects should be equidistant from this initial fixation position. In order to reduce the likelihood of subjects inferring the experimental predictions due to repeated exposure to particular patterns or relationships in the stimuli, filler objects, as well as filler trials, are recommended. Highly complex displays, such as photographs that have objects partially occluding other objects in depth, or objects whose adjoined and abutting parts are important for separate analysis, can prove rather difficult for eye-position coding. Especially if data analysis is being performed by trained coders watching frame-by-frame videotape, but even when analysis is automated in x,y computer screen coordinates, just a small amount of noise or jitter in the eye position signal can introduce uncertainty in whether one abutting object/part or the other is actually being fixated. A common solution for this is to arrange the visual displays such that there is white space and/or a contour-based divider separating each relevant region or object by at least a couple degrees of visual angle.

6.4 "Blind" coding

If the eye-position record is being analyzed by human coders via frame-by-frame videotape, it is, of course, wise for the coders to be prevented from knowing the experimental conditions and predictions for each trial. If the critical experimental manipulation is in the auditory portion of the videotape, this can sometimes be solved by simply coding the silent video portion. However, in other circumstances, trained coders who are unaware of the experimental manipulation and the theoretical predictions may need to analyze each trial with a spreadsheet that uses coded labels for conditions. A more recent solution to this problem, is to store eye-position as x,y screen coordinates and map them onto the x,y coordinates of objects on the same screen. Current headband-mounted eye-tracking systems allow this automated data analysis as long as the visual display is presented on a computer screen and the subject makes rather minimal head and trunk movements in front of that screen.

6.5 Participant ease

Whether it is a bitebar on which the subject must make a form-fitting dental impression so that her head will be held in place during the experiment or a 3/4-pound headband that the subject is being asked to wear, eye-tracking equipment looks and feels intimidating to a newcomer. And experimental findings from uptight and uncomfortable participants may not generalize to how they behave in normal everyday settings. Therefore, a good tactic when the participant first enters the lab and begins filling out the consent form, is to chat with them a bit about the classes they're taking etc., and then describe the basics of how the eyetracker works, while demonstrating it on yourself. This way they get to see someone else wearing the equipment with ease. Also, wearing a headband-mounted eyetracker for more than half an hour can sometimes cause a headache. Experiments that last longer than 30 minutes should probably introduce a 10-minute intermission.

6.6 Practice

Every eye-tracking system has its own set of unique tricks and parameter settings that take time to learn. In particular, achieving an accurate calibration (usually a 5–10 minute process), such that the subject's actual eye position is correctly indicated for all regions of the display, is something that requires careful attention to parameters such as centering the eye in the eye camera's view, reducing distracting reflections on the sclera, as well as a certain amount of speed and fluidity in entering data for the calibration positions to minimize head-drift during calibration. The typical graduate student can expect to require a couple weeks of practice before being able to complete a good calibration for the majority of their experimental subjects. And there will always be some portion of experimental subjects, 5–10%, for whom three or even four attempts at calibration simply fail to produce an accurate record of eye-position. This can be due to a variety of things, such as very light or very dark irises (with some trackers), naturally droopy eyelids, downward-pointing eyelashes that obstruct the eye camera's view, the headband not fitting the person's head, or even incorrigible head motion during the calibration phase.

It could be said that the actual tracking of eye movements is more of an art than a science. And there are certainly dozens of minor tricks that a research team will learn and develop through practice with any given eyetracking apparatus and experimental setup that cannot be anticipated in advance. However, the small handful of practical tips provided in this section are unlikely to be found anywhere else in the literature. We hope they prove helpful.

7. Conclusion

We contend that eye movements provide an index of real-time mental activity that most other methodologies do not (but cf. event-related potentials, Coulson, this volume, and continuous kinematic properties of manual responses, Abrams & Balota 1991; Coles, Gratton, Bashore, Eriksen, & Donchin 1985; Spivey, Grosjean, & Knoblich 2005). Eye movements provide a semi-continuous record of the time-course of partially-active representations competing for an overt skeletal motor response, and therefore offer rich insight into *how cognition happens* – not just the outcomes it produces. In Cognitive Linguistics' concern with the wide array of perceptual, cognitive, and motor faculties that underlie language use, it stands to gain considerably from eye movement techniques.

Acknowledgements

The authors are grateful to four supportive and constructive anonymous reviewers for providing comments that improved the original version of the manuscript. Work on this chapter was supported by NIMH grant R01-63961 to the third author.

References

Abrams, R. A. & Balota, D. A. (1991). Mental chronometry: Beyond reaction time. *Psychological Science, 2*(3), 153–157.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language. 38*(4), 419–439.

Altmann, G. T. M. & Kamide, Y. (2004). Now you see it, now you don't: mediating the mapping between language and the visual world. In J. M. Henderson & F. Ferreira (Eds.), *Interfacing Language, Vision, and Action*. San Diego: Academic Press.

Altmann, G. T. M. & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30*(3), 191–238.

Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science, 15*(4), 225–231.

Antrobus, J. S. & Antrobus, J. S. (1969). Rapid eye movements and rapid eye movement periods. *Psychophysiology, 6*(1), 45–48.

Baddeley, A. (1986). *Working memory*. New York, NY, US: Clarendon Press/Oxford University Press.

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*(1), 66–80.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 2*, 723–767.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660.

Brandt, S. A. & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience, 9*(1), 27–38.

Bridgeman, B. (1992). Conscious vs unconscious processes: The case of vision. *Theory and Psychology, 2*(1), 73–88.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art*. Chicago, IL: Univ. Chicago Press.

Carpenter, P. A. & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). New York: Academic Press.

Chafee, M. V. & Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memory task. *Journal of Neurophysiology, 79*(6), 2919–2940.

Chafee, M. V. & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of Neurophysiology, 83*(3), 1550–1566.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 687–696.

Charness, N. Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory and Cognition*, *29*(8), 1146–1152.

Clark, H. (1916). Visual imagery and attention: An analytical study. *American Journal of Psychology, 27*(4), 461–492.

Coles, M., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 529–553.

Corbetta, M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Ollinger, J. M., Drury, H. A., et al. (1998). A common network of functional areas for attention and eye movements. *Neuron, 21*, 761–773.

Cornsweet, T. N. & Crane, H. D. (1973). Accurate two-dimensional eye tracker using first and fourth Purkinje images. *Journal of the Optical Society of America, 63*(8), 921–928.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33*(1–2), 25–62.

Delabarre, E. B. (1898). A method of recording eye-movements. *American Journal of Psychology, 9*(4), 572–574.

Demarais, A. M. & Cohen, B. H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology, 49*(3), 229–247.

Diefendorf, A. R. & Dodge, R. (1908). An Experimental Study of the Ocular Reactions of the Insane from Photographic Records. *Brain, 31*, 451–492.

Duhamel, J., Colby, C., & Goldberg, M. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*, 90–92.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*(6), 409–436.

Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language, 27*(2), 143–165.

Fernald, R. D. (1997). The evolution of eyes. *Brain Behav Evol, 50*(4), 253–259.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348–368.

Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*, 516–547.

Findlay, J. M. & Gilchrist, I. D. (1998). Eye guidance and visual search. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 295–312). Oxford, England: Elsevier Science Ltd.

Goldman-Rakic, P. S., Chafee, M., & Friedman, H. (1993). Allocation of function in distributed circuits. In T. Ono, L. R. Squire & et al. (Eds.), *Brain mechanisms of perception and memory: From neuron to behavior* (pp. 445–456). London: Oxford University Press.

Goldthwait, C. (1933). Relation of eye movements to visual imagery. *American Journal of Psychology, 45*, 106–110.

Goodenough, D. R., Shapiro, A., Holden, M., & Steinschriber, L. (1959). A comparison of "dreamers" and "nondreamers": Eye movements, electroencephalograms, and the recall of dreams. *Journal of Abnormal and Social Psychology, 59*, 295–302.

Griffin, Z. & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274–279.

Grant, E. R. & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*(5), 462–466.

Hale, S. M. & Simpson, H. M. (1971). Effects of eye movements on the rate of discovery and the vividness of visual images. *Perception & Psychophysics, 9*(2–B), 242–246.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision Research, 38*(1), 125–137.

Hebb, D. O. (1968). Concerning imagery. *Psychological Review, 75*, 466–477.

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*(5), 1084–1102.

Hegarty, M. & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory & Language, 32*(6), 717–742.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences, 7*, 498–504.

Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience, 12*(5), 894–907.

Hoffman, J. E. & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception and Psychophysics, 57*(6), 787–795.

Huey, E. B. (1898). Preliminary Experiments in the Physiology and Psychology of Reading. *American Journal of Psychology, 9*(4), 575–886.

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 29*(5), 1017–1027.

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition, 29*(7), 1000–1009.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition, 95*(1), 95–127.

Kosslyn, S. M., Behrmann, M., & Jeannerod, M. (1995). The cognitive neuroscience of mental imagery. *Neuropsychologia, 33*(11), 1335–1344.

Laeng, B. & Teodorescu, D.-S. (2002). Eye scanpaths during visual imagery reenact those of perception of the same visual scene. *Cognitive Science, 26*(2), 207–231.

Land, M. F. (1995). The functions of eye movements in animals remote from man. In J. M. Findlay & R. Walker (Eds.), *Eye movement research: Mechanisms, processes and applications. Studies in visual information processing, 6* (pp. 63–76). New York, NY: Elsevier Science.

Land, M. F. & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25–26), 3559–3565.

Land, M. F. & Lee, D. N. (1994). Where do we look when we steer. *Nature, 369*(6483), 742–744.

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*(7), 963–977.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358–368.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science, 27*(5), 801–805.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227.

Marian, V. & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193.

Martin, A. (2001). Functional neuroimaging of semantic memory. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional NeuroImaging of Cognition* (pp. 153–186). Cambridge: MIT Press.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899–917.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition, 32*, 1389–1400.

Matlock, T. & Richardson, D. C. (2004). Do eye movements go with fictive motion? *Proceedings of the Twenty-sixth Annual Meeting of the Cognitive Science Society.*

McClelland, J. L. & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1–86.

McMurray, B., Tannenhaus, M. K., Aslin, R. N., & Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. *Journal of Psycholinguistic Research, 32*(1), 77–97.

Merchant, J., Morrissette, R., & Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Transactions on Biomedical Engineering, 21*(4), 309–317.

Moore, T. & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature, 421*(6921), 370–373.

Moore, T. & Fallah, M. (2001). Control of eye movements and spatial attention. *Proceedings of the National Academy of Science, USA, 98*(3), 1273–1276.

Morrison, R. E. & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. *Perception and Psychophysics, 30*(4), 395–396.

Murray, W. S. & Kennedy, A. (1988). Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 40*(4–A), 693–718.

Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT, US: Appleton-Century-Crofts.

O'Regan, J. K. (1992). Solving the 'real' mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–488.

Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology, 21*, 422–452.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3–25.

Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: a sketch of the FINST spatial index model. *Cognition, 32*, 65–97.

Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358–374.

Rayner, K. & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision Research, 16*(8), 829–837.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & et al. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4*(3–4), 21–49.

Richardson, D. C. & Kirkham, N. Z. (2004). Multimodal Events and Moving Locations: Eye Movements of Adults and 6-Month-Olds Reveal Dynamic Spatial Indexing. *Journal of Experimental Psychology: General, 133*(1), 46–62.

Richardson, D. C. & Spivey, M. J. (2000). Representation, space and Hollywood Squares: Looking at things that aren't there anymore. *Cognition, 76*, 269–295.

Richardson, D. C. & Spivey, M. J. (2004a). Eye tracking: Characteristics and methods. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering*. Marcel Dekker, Inc.

Richardson, D. C. & Spivey, M. J. (2004b). Eye tracking: Research areas and applications. In G. Wnek & G. Bowlin (Eds.), *Encyclopedia of Biomaterials and Biomedical Engineering*. Marcel Dekker, Inc.

Robinson, D. A. & Fuchs, A. F. (1969). Eye movements evoked by stimulation of frontal eye fields. *Journal of Neurophysiology, 32*(5), 637–648.

Roffwarg, H. P., Dement, W. C., Muzio, J. N., & Fisher, C. (1962). Dream imagery: Relationship to rapid eye movements of sleep. *Archives of General Psychiatry, 7*(4), 235–258.

Rozenblit, L., Spivey, M. J., & Wojslawowicz, J. (2002). Mechanical Reasoning about Gear-and-belt Diagrams: Do Eye-movements Predict Performance? In M. Anderson, B. Meyer, & P. Olivier (Eds.), *Diagrammatic Representation and Reasoning* (pp. 223–240). Berlin: Springer-Verlag.

Ruggieri, V. (1999). The running horse stops: The hypothetical role of the eyes in imagery of movement. *Perceptual & Motor Skills, 89*(3, Pt 2), 1088–1092.

Ryle, G. (1949). *The concept of mind*. New York: Barnes and Noble.

Sheliga, B., Riggio, L., Craighero, L., & Rizzolatti, G. (1995). Spatial attention-determined modifications in saccade trajectories. *Neuroreport: An International Journal for the Rapid Communication of Research in Neuroscience, 6*(3), 585–588.

Sheliga, B. M., Riggio, L., & Rizzolatti, G. (1994). Orienting of attention and eye movements. *Experimental Brain Research, 98*, 507–522.

Shepherd, M., Findlay, J. M., & Hockey, R. J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 38*(3–A), 475–491.

Spivey, M. J. & Geng, J. J. (2001). Oculomotor mechanisms activated by imagery and memory: Eye movements to absent objects. *Psychological Research/Psychologische Forschung, 65*(4), 235–241.

Spivey, M., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences*, *102*, 10393–10398.

Spivey, M. J. & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science, 10*(3), 281–284.

Spivey, M. J., Richardson, D. C., & Fitneva, S. (2004). Thinking outside the brain: oculomotor indexes to visual and linguistic information. In J. M. Henderson & F. Ferreira (Eds.), *Interfacing Language, Vision, and Action* (pp. 161–190). San Diego: Academic Press.

Spivey, M. J. & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 24*(6), 1521–1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*(4), 447–481.

Spivey, M. J., Tyler, M., Richardson, D. C., & Young, E. (2000). *Eye movements during comprehension of spoken scene descriptions.* Paper presented at the 22nd Annual Conference of the Cognitive Science Society, Philadelphia.

Spivey-Knowlton, M. & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition, 55*(3), 227–267.

Spivey-Knowlton, M., Tanenhaus, M., Eberhard, K., & Sedivy, J. (1998). Integration of visuospatial and linguistic information in real-time and real-space. In P. Olivier & K. Gapp (Eds.), *Representation and Processing of Spatial Expressions* (pp. 201–214). Mahwah, NJ: Erlbaum.

Stoy, E. G. (1930). A preliminary study of ocular attitudes in thinking of spatial relations. *Journal of General Psychology, 4*, 379–385.

Stratton, G. M. (1906). Symmetry, linear illusion, and the movements of the eye. *Psychological Review, 13*(2), 82–96.

Tanenhaus, M. K., Spivey Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632–1634.

Tanenhaus, M. K. & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language & Cognitive Processes, 11*(6), 583–588.

Tanenhaus, M. K. & Trueswell, J. C. (1995). Sentence comprehension. In J. L. Miller & P. D. Eimas (Eds.), *Speech, language, and communication Vol. 11*, (pp. 217–262). San Diego, CA, US: Academic Press.

Thiele, A., Henning, P., Kubischik, M., & Hoffmann, K. P. (2002). Neural mechanisms of saccadic suppression. *Science, 295*(5564), 2460–2462.

Tinker, M. A. (1928). A photographic study of eye movements in reading formulae. *Genetic Psychology Monographs, 3*(2), 68–182.

Totten, E. (1935). Eye movement during visual imagery. *Comparative Psychology Monographs, 11*, 46.

Treue, S. (2001). Neural correlates of attention in primate visual cortex. *Trends in Neurosciences, 24*, 295–300.

Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments and Computers, 34*(4), 518–528.

Wooding, D. S., Muggelstone, M. D., Purdy, K. J., & Gale, A. G. (2002). Eye movements of large populations: I. Implementation and performance of an autonomous public eye tracker. *Behavior Research Methods, Instruments and Computers, 34*(4), 509–517.

Yarbus, A. L. (1965). *Role of eye movements in the visual process*. Oxford, England: Nauka.

Young, L. R. (1970). Recording Eye Position. In M. Clynes & J. H. Milsom (Eds.), *Biomedical Engineering Systems*. New York: McGraw-Hill.

Speaking for the wordless

Methods for studying the foundations of cognitive linguistics in infants

Amanda Brandone, Roberta Michnick Golinkoff,
Rachel Pulverman, Mandy J. Maguire, Kathy Hirsh-Pasek,
and Shannon M. Pruden

1. Introduction

Cognitive linguistics seeks to characterize language in relation to more general cognitive processes, to give an account of the relationship between language and thought, and to uncover the cognitive foundations of language. Research with adults has been the primary vehicle for validating the theoretical assertions of cognitive linguistics, as papers presented in this volume attest. However, research with the youngest of participants – infants – offers an exciting new perspective for the field of cognitive linguistics. Through experimentation with preverbal infants it is possible to study the very origins of the cognitive linguistic concepts that underlie language.

Long before any language has been acquired, infants are developing concepts. According to Mandler (1991), many of these prelinguistic conceptual primitives – the foundations of the human conceptual system – are well-established by the time language appears. For example, Mandler (2006) argues that, in addition to their knowledge of objects:

> By the end of the first year infants have an extensive repertoire of relational concepts that provides enough understanding to learn a host of motion verbs and prepositions. Infants understand that actions are goal-directed, have beginnings, middles, and ends, and are often linked to other actions. They conceptualize actions as going into and out of things, behind objects, onto surfaces, up and down, joining and separating, and fitting tightly or loosely. They also conceptualize the difference between causal and noncausal actions and the difference between agents and recipients of action. (p. 128)

Research into the nature and representational format of these preverbal concepts in infancy can act as a bridge in understanding the remarkable transition from nonlinguistic cognition to language. In essence, research with preverbal infants allows researchers to get to the heart of cognitive linguistics, revealing our initial, universal conceptual foundation and how it merges with and is shaped by language throughout development.

How can researchers get inside the minds of children who have no way of expressing their cognitive, let alone their linguistic, ideas? Is it even possible to study cognitive linguistic concepts in preverbal infants? In this chapter, we will describe two research methods that can be used to reveal the fundamental prelinguistic concepts of young infants. These methods are *habituation* and the *intermodal preferential looking paradigm* (IPLP). After illustrating these methodologies we will then describe experimental findings from our own laboratories that demonstrate the contribution these paradigms are beginning to make to cognitive linguistics in the area of verb learning.

2. The habituation paradigm

2.1 Rationale

Long before researchers can expect infants to answer questions or even follow directions, researchers must utilize simple, yet devious methods that make use of the behaviors that are within the repertoire of the preverbal infant. One such clear, experimentally manipulable behavior is infants' visual fixation. If you place an infant in a neutral environment and then initiate a stimulus event, the infant will naturally visually orient and attend to that stimulus. Upon repeated exposure to the same stimulus, however, the infant will eventually lose interest and his or her visual attention to the original stimulus will diminish (Bornstein 1985).[1] This most basic, involuntary infant behavior has long been capitalized on in the habituation paradigm. In this section, we will provide a brief history, rationale, and general explanation of the habituation paradigm. We will then move to a more specific illustration of the habituation procedure, followed by a justification of the relevance of habituation to the field of cognitive linguistics.

The habituation paradigm grew from the work of Robert Fantz in the late 1950s and early 1960s. In his research on infant vision, Fantz was among the first to make use of the phenomenon of habituation. He observed that babies systematically look at different visual stimuli for varying lengths of time; that is, infants look selectively at a preferred visual stimulus (Fantz 1964). This observation provided researchers with a simple, noninvasive method for accessing the cognitive processes of preverbal infants: By measuring how much more babies look at one display than at another, researchers can determine what infants can see and discriminate in the visual display.

Typically, in the visual habituation paradigm, infants as young as 2 months of age are presented repeatedly with the same stimulus or stimulus set until they demonstrate a lack of interest by attending less and less. Following habituation, a new stimulus set is presented. If infants do not discriminate between the new stimulus and the habituated stimulus, they should look at the new stimulus equally or less than at the previously viewed displays. However, if they do discriminate between the new and the habituated stimuli, infants should show renewed interest, manifested by *increased* looking and prolonged at-

1. See Sokolov (1963) and Groves and Thompson (1970) for descriptions of two prominent theories of the mechanisms responsible for the habituation process.

tention to the new stimulus. Through careful manipulation of test stimuli, researchers can close in on the nature of the difference to which the infant is sensitive.

Take as a simple example the question of infants' ability to discriminate objects on the basis of shape. In order to evaluate this question, during habituation, infants could be shown a series of rectangles of varying sizes and colors. Once they had decreased their looking to these objects, the experimenter could present a different shape, for example, a triangle. If infants *are* able to discriminate between a rectangle and a triangle, they should *increase* their looking to the novel triangle as compared to a novel rectangle. However, if infants cannot discriminate between these shapes, they should remain habituated and visual fixation to the triangle should not increase. Systematic results from this simple experimental design would allow researchers to conclude, at the very least, whether infants are able to make a distinction between examples of two basic geometric shapes.

In the decades since Fantz' seminal discovery, the habituation paradigm has been developed and fine-tuned so that it can be used in countless ways to document the cognitive processes of preverbal infants. Throughout its lifespan, habituation has been applied to questions of perceptual detection and discrimination (of color, (e.g., Bornstein 1981); of forms, (e.g., Cornell 1975); of faces, (e.g., Barrera & Maurer 1981)), categorization (e.g., Bornstein 1981), memory (e.g., Lasky & Spiro 1980), and concept formation (e.g., Ross 1980), and to the assessment of individual differences in cognition (e.g., Bornstein 1985). Of late, it has also been used to reveal abilities relevant to the study of cognitive linguistics, including infants' ability to extract spatial relations, such as containment, support, tight-fit (Casasola & Cohen 2002), above, below (Quinn, Cummins, Kase, Martin, & Weisman 1996), and between (Quinn, Adams, Kennedy, Shettler, & Wasnik 2003), and motion event components, such as path and manner (Pulverman 2005; Pulverman & Golinkoff 2003; Pulverman, Golinkoff, Hirsh-Pasek, & Sootsman Buresh under review).

2.2 Procedure

Although the design of the habituation paradigm is fairly straightforward, the exact experimental procedures vary according to the empirical question under scrutiny and the research facilities. Despite this variation, by definition the habituation paradigm requires that infants are familiarized with a series of experimental stimuli. These stimuli can be sequences of video (e.g., Pulverman 2005), still images (e.g., Xu & Spelke 2000), and/or live action events (e.g., Woodward 1998), either with or without accompanying linguistic input. The specific method of presentation of these stimuli varies depending on the habituation design. Although there are a variety of habituation regimens, each with its own unique attributes and contributions (see Bornstein 1985), the crucial distinction between habituation designs lies in whether the experiment is *infant-* or *experimenter-*controlled.

In experimenter-controlled procedures, stimulus presentation is guided by predetermined experimental parameters, such as trial length or number of trials. Infants are presented with a stimulus a fixed number of times for a fixed duration, irrespective of the infant's behavior during the experiment. In such a procedure, the transition from the habituation to the test phase of the experiment is also predetermined. Infants' visual fixation during each presentation of a stimulus is recorded and analyzed, but, crucially, has no effect on the course of the experiment.

In contrast, infant-controlled procedures put the infant in charge of the stimulus presentation. In an infant-controlled habituation design, stimuli are presented to the infant as determined by the child's visual behavior during the course of the experiment. Each trial is ended either when the child looks away from the stimulus for a predetermined amount of time (typically approximately two seconds), or when the trial has reached a set, maximum duration (e.g., 30 seconds). An attention-getting stimulus is often presented at the end of a trial in an infant-controlled design in order to reorient the child to the visual display. As soon as the child has fixated on the stimulus again, a new trial begins.

A trained observer, blind to the habituation and test stimuli being displayed, observes the infant on-line during the course of an infant-controlled experiment. The observer records and times the duration of the infant's visual fixation to the stimulus using a special computer program (e.g., Cohen, Atkinson, & Chaput 2000). Infants are repeatedly presented with the same stimulus or stimulus set until their visual fixation time to the stimulus decreases by a predetermined amount, relative to some previous, higher level of fixation observed during the experimental session. For example, in our laboratory, the habituation criterion is defined by averaging looking times during the first three habituation trials. The end of the habituation phase is signaled when the infant's average looking time across three consecutive trials drops to 50% or less of the mean of the first three trials.[2] Once this decrement has been attained, test stimuli are presented.

Test stimuli are carefully selected on the basis of the type of experimental question being asked. Test events typically include a presentation of something *different* from the habituation stimuli and something the *same* (the no-change control). Visual fixation to a new stimulus is recorded as a measure of the infant's renewed interest or continued habituation to the stimulus display. The sequence of the placement of test trials, including the control trial, is counterbalanced across subjects. Immediately following the test phase a highly attractive recovery stimulus is often presented to ensure that participants are not too fatigued to attend to experimental stimuli at the conclusion of the study. In our own laboratory, a video of a laughing baby is used for the recovery trial. A criterion is set such that if the infant does not dishabituate to the recovery stimulus the child's data may be discarded due to fatigue.

Following the experimental procedure, standard statistical analyses are used to compare infants' performance across habituation and test trials. Typical between-subjects independent variables include age, gender, and linguistic level. The dependent variable is always the infant's visual fixation time.

---

2. This is just one example of a common criterion of habituation. The exact criterion can be manipulated by the computer program depending on the age of the participants, the nature of the experimental stimuli, etc.

2.3 Relevance of habituation to the study of cognitive linguistics

While the observed behavior in a habituation experiment is merely a waning and waxing of expressed interest made manifest through looking behavior, researchers infer much more on the part of an infant during habituation and dishabituation. The child's looking behavior is believed to reflect reciprocal mental processes involving the construction of and comparison to a mental representation of the stimulus display. During the course of habituation, the infant is thought to process and encode information in the stimulus display and to then internalize that information in the form of a mental representation. As they do so, visual fixation declines. In the test portion of the experiment, infants compare each new stimulus with their mental representation of the original display. Increased looking to a subtly different manifestation of the original stimulus suggests that infants are able to detect and encode that difference, thus discriminating between the stimuli. Generalization of the habituation stimulus to the test stimulus, however, reveals that the infant's representation of the new display is functionally or conceptually equivalent to that of the original display: The key distinction either has not been encoded or is not considered significant by the infant.

How does this method inform the study of cognitive linguistics? In order for an infant to mentally represent a stimulus event, the infant must possess some means of conceptualizing that event. Detecting a change in stimulus events thus suggests that the infant is able to encode those events and the crucial distinction between them.[3] Through careful manipulation of stimulus events, experimenters can thus demonstrate the existence of a prelinguistic concept that is the foundation for a subsequent linguistic one.

Consider, for example, the study of spatial cognition and language-specific semantic spatial categories. Languages differ greatly in the way they semantically categorize spatial relations (e.g., Bowerman 1996; Bowerman & Choi 2001). For example, English makes a distinction between actions resulting in containment (*put in*, e.g., books in a bag, finger in a ring) versus those resulting in support or surface attachment (*put on*, e.g., bowl on a table, glove on a hand, magnet on a refrigerator), whereas Korean makes a cross-cutting distinction between tight-fit relations (*kkita*, e.g., finger in a ring, pieces in a puzzle, glove on a hand, cap on a pen) versus loose-fit or other contact relations (various verbs, e.g., books in a bag, bowl on a table, magnet on a refrigerator). In particular, the Korean verb *kkita* refers to actions resulting in a tight-fit relation regardless of containment or support. These cross-linguistic differences raise the following question: Do all infants, regardless of linguistic environment, display a sensitivity to the spatial relations of containment, support, tight-fit and loose-fit? To address a piece of this question, in a between-subjects design, Spelke and Hespos (2002) habituated 5-month-old English-learning infants to events in which an object was placed in either a tight-fitting or a loose-fitting containment relation. Following habituation, infants were shown both the

3. The discrimination of physical/perceptual differences does not necessarily require conceptual knowledge (e.g., Quinn, Adams, Kennedy, Shettler, & Wasnik 2003). Caution must be taken in experimentation to ensure that a perceptual change representative of the abstract concept is not detected independently of conceptual meaning.

relation from habituation and the new relation. Results indicated that, by 5-months of age, English-learning infants successfully discriminated between placing an object in a tight-fitting containment relation and placing an object in a loose-fitting containment relation. These results lend support to the conclusion that infants possess prelinguistic concepts of spatial relations, including those that are not lexically encoded in their language (see also McDonough, Choi, & Mandler 2003; Quinn et al. 1996; Quinn et al. 2003). As this example demonstrates, by applying the habituation paradigm to the field of cognitive linguistics, researchers have access to a tool capable of revealing the finite set of cognitive linguistic concepts with which the preverbal infant is equipped.

3. The intermodal preferential looking paradigm

3.1 Rationale

Like the habituation paradigm, the intermodal preferential looking paradigm (Hirsh-Pasek & Golinkoff 1996a), or IPLP, also uses visual fixation as the dependent measure. However, instead of manipulating novelty, the IPLP asks whether infants can use their early understanding of language to find the correspondence or link between stimuli presented in the visual and auditory modalities (Figure 1). Next we provide a brief history, rationale, and general explanation of the IPLP, followed by a more specific description of the IPLP procedure, and a justification of its application to the study of cognitive linguistics.

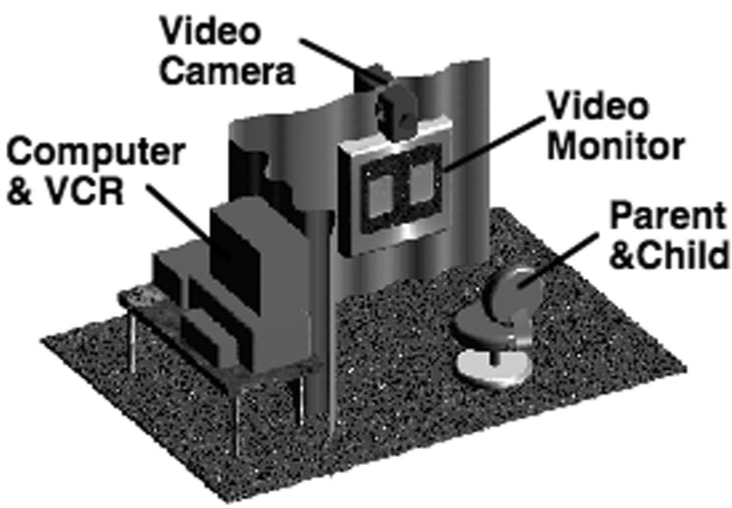


Figure 1. Intermodal preferential looking paradigm

The IPLP was adapted from the work of Spelke (1979), who developed it to study intermodal perception. In Spelke's version, 4-month-old infants saw two events (e.g., a person clapping her hands and a donkey falling onto a table) and heard a nonlinguistic auditory stimulus that matched only one of the events (e.g., the sound of hands clapping). Infants tended to look more at the screen on which the event matched the auditory stimulus than at the screen on which it did not. The IPLP represents an extension of the Spelke paradigm into a new realm. It is among the first to demonstrate that infants will use a symbolic stimulus – language – to guide their search for a matching event.

In our own adaptation of the paradigm (Golinkoff, Hirsh-Pasek, Cauley, & Gordon 1987; Hirsh-Pasek & Golinkoff 1996a), an infant is presented with two side-by-side video images. An audio speaker plays a linguistic stimulus that is consonant with, or matches, only one of the pictured displays. The infant's task is to look at one of the two video images. In all such studies the logic is the same as in Spelke's experiment: Infants should allocate more attention to the video event that matches what they are hearing (in this case, a linguistic message) than to a video event that does not match what they are hearing. In this linguistic version of the preferential looking paradigm, children can show relatively more visual attention to the match than to the nonmatch only if they understand the language that is used. This methodology is ideal for young children and infants as young as 7 months of age because it places minimal demands on the child: In making decisions about how language maps to events, infants need only move their eyes.

To see how the paradigm works, consider a pair of stimuli in what is arguably the simplest case: noun comprehension (Golinkoff et al. 1987). In one trial, a shoe appeared on one screen and a boat on the other. The linguistic message (produced in child-directed speech) was "Where's the shoe? Find the shoe!" The hypothesis, which was confirmed, was that infants would look longer toward the screen displaying the shoe than toward the screen displaying the boat. Using this basic premise, the IPLP can be applied to innumerable research questions, such as infants' sensitivity to the cues for constituent structure (Hirsh-Pasek & Golinkoff 1996a, b), to the meaning of common nouns and verbs (Hirsh-Pasek & Golinkoff 1996a, b), to word order (Hirsh-Pasek & Golinkoff 1996a, b), to morphological word endings (Golinkoff, Hirsh-Pasek, & Schweisguth 2001), and to information about sentence frames (Bavin, Wales, & Kelly 1998; Hirsh-Pasek & Golinkoff 1991, 1993; Naigles 1990; see Hirsh-Pasek & Golinkoff 1996b for a review).

3.2 Procedure

In order to illustrate more fully the specific design of a cognitive linguistic experiment using the IPLP, we will describe the setup and procedure used in our own laboratory. The infant is seated on a parent's lap directly in front of a large television. Black curtains that obscure the child's view of the experimental apparatus and block any possible distraction surround the television screen. A concealed video camera records the infant's face while a hidden VCR displays the video stimuli. Parents are instructed to hold their child firmly around the waist, to close their eyes, and to refrain from speaking or directing their child's attention toward or away from the screen during the study.

Two videos are developed for each study. The videos are constructed so that they work in synchrony, trial for trial, down to the same number of frames. Videos are then edited such that they may be played simultaneously, side by side, on the television screen. Although achieving perfect control is impossible, every effort is made to balance the amount of action and color on each side of the screen so as to reduce any potential salience problems.

The exact video design varies depending on the research question; however, cognitive linguistic experiments in the IPLP typically involve at least three phases: *salience*, *training*, and *test*. A centering stimulus is played between each training and test clip in order to both reorient the infant's attention to the center of the screen and maintain the infant's interest throughout the experiment. In our laboratories, video of a laughing baby's face in the center of the television screen is used. The centering stimulus is accompanied by audio directing the infant's attention to the screen and/or introducing the subsequent video event.

3.2.1 *Salience phase*

The purpose of the salience phase is to test whether either of the images to be presented later in the test phase is inherently more interesting than the other, prior to training or test exposure. Thus, in the salience phase, test images are presented side by side accompanied by neutral audio that is consistent with both events. If the visual stimuli have been well balanced for perceptual factors, while the linguistic stimulus is still neutral, it is predicted that across participants attention should be distributed approximately equally to each member of the simultaneous pair: Infants, overall, should not prefer one stimulus over another.

3.2.2 *Training or sequential trials*

In a word-learning study, during the training phase participants are given a novel label for an object or event displayed on the screen. For example, infants might see a girl performing a novel action and hear, "Look! She's *daxing*! Do you see her *daxing*?" Often, multiple training trials are shown sequentially, separated by the centering stimulus.

In a syntactic study, sequential trials serve to introduce the video events before the child is asked to find the match for the linguistic stimulus. Children see the video events that will later be used in test, sequentially and individually, accompanied by neutral audio, to familiarize them with the test events. For example, infants might see video of a woman kissing a ball while dangling keys in the foreground, followed sequentially by video of a woman kissing keys while holding a ball in the foreground (see Hirsh-Pasek & Golinkoff 1996b). The order of the stimulus presentation is counterbalanced across participants.

3.2.3 *Test trials*

The test trials are identical to the salience trials in all ways but one: The linguistic stimulus that now accompanies these trials exhorts the child to look at the screen that matches the linguistic stimulus. For example, in the syntactic study described above, infants would see the video of the woman kissing the ball with keys in the foreground next to the video of the woman kissing the keys with a ball in the foreground, and hear, "Where is she *kissing* *the keys*?" In the word-learning study described above, infants would see the previously labeled novel action next to a new, novel action (seen earlier in the salience phase), and hear, "Where is she *daxing?*" The location of test images on the screen is counterbalanced across subjects. In studies of word learning, initial trials affording the primary test of language learning occur in pairs separated by the centering stimulus. Often a *new label* and a *recovery* test follow the *initial* test trials as a strict, additional examination of word learning. In the *new label* test, infants see the same visual stimuli as in the test trials, however the accompanying linguistic stimulus is now inconsistent with the original, target, visual stimulus (e.g., "Where is she *clooming?*"). Here, looking time to the initial clip should decrease. In the *recovery* phase, infants are again shown the images from the initial test phase accompanied by the original linguistic stimulus (e.g., "Where is she *daxing?*"). The proportion of looking time to the initial clip should now increase as compared to the *new label* test trial.

The linguistic stimulus for each test trial is presented first during the centering trial that precedes the test trial, and then again during the test trial itself. The linguistic stimulus lasts for the duration of the test event to maximize the amount of time the infant looks toward the matching screen. In our laboratories, all studies use child-directed speech produced by a female, native speaker of the language. Despite the added linguistic complexity, in order to make the audio more interesting and more natural, all linguistic stimuli are presented in the form of full sentences.

Infants' visual fixation throughout the study is recorded by a concealed video camera. After the completion of the experiment, the videotape is reviewed and coded offline by an experienced coder, blind to the condition the child is seeing. The coder watches the child and records to the hundredth of a second the amount of time the child looks toward specific areas of the screen (right, left, and center). In our laboratories, it is standard procedure for a coder to code each trial at least twice or more until visual fixation times for each trial are within 0.3 seconds of each other. In addition, a different reliable coder codes a certain percentage of the subjects again, in order to ensure inter-rater reliability.

Following the offline coding, standard statistical analyses are used. Typical between-subjects independent variables in these studies are age, gender, and linguistic level. Visual fixation time – the total amount of time in seconds the child spends watching the matching versus the nonmatching screen – is the dependent variable.[4]

The hypothesis in each IPLP study is that significantly more visual fixation time will be allocated to the matching than to the nonmatching event. Note that it is never hypothesized that the child will look exclusively at the matching screen. In fact, looking times are expected to be distributed between the video events since, for control purposes, visual displays in a pair are designed, as much as possible, to contain equivalent activity, bright colors, and complexity.

---

4. After the age of 3 years, IPLP looking procedures can be replaced by pointing procedures. Using this method, children are asked to point to the side of the screen that is consistent with the linguistic input. For example, children might be asked to "Point to *blicking!*" rather than simply to "Look at *blicking!*"

3.3 Relevance of the IPLP to the study of cognitive linguistics

Studies of language comprehension occurring prior to advanced language production offer researchers a unique view of the emergent language system and the maturing language learner. By employing a natural response already in the repertoire of children in the early stages of language development, the IPLP provides a means of investigating children's linguistic capabilities before total indoctrination into their native language. Such a tool also allows researchers to probe the interaction of underlying concepts and language input at its earliest stages.

As an example, consider once again the study of spatial cognition and language-specific spatial categories. Using the IPLP, Choi, McDonough, Bowerman, and Mandler (1999) assessed English- and Korean-learning toddlers' comprehension of *in* (containment) and *kkita* (tight-fit), respectively. Participants watched pairs of scenes of spatial relations (e.g., putting pegs into holes in a wooden block vs. putting pegs on top of the wooden block; putting rings loosely into a basket vs. putting rings tightly onto a tapered pole) while listening to sentences with and without the target word. Results revealed that by 18 months, Korean-learning toddlers comprehended *kkita* as referring to both tight-fit containment and tight-fit support events, whereas English-learning toddlers comprehended *in* as referring to both tight-fit and loose-fit containment events. These results suggest that, despite their universal starting point, before their second birthdays, English- and Korean-learning toddlers have begun to acquire the semantic spatial categories specific to their language: Toddlers attend to the spatial distinctions that are linguistically relevant and disregard those that are not. As this example illustrates, the IPLP offers researchers a useful method of studying what happens when presumably universal, prelinguistic concepts meet the influence of a specific language.

4. Verb learning: A case study

In order to illustrate more fully how habituation and the intermodal preferential looking paradigm can inform the cognitive linguistic program, we will illustrate with a case study from our own laboratories. The research we will describe is an attempt to use the cognitive linguistic perspective to address the complexity of the process of verb learning. We will first discuss the theoretical problem of verb learning from a cognitive linguistic perspective and then shed light on that problem using evidence from the habituation and intermodal preferential looking paradigms.

4.1 The problem of verb learning

Verbs are the architectural centerpiece of a sentence. They allow us to talk about not only people and things, but, crucially, the relations between them. Verbs are necessary to express everything from the simplest to the most complex events; without a verb, there is no sentence. However verb learning has proven to be a difficult task for children (e.g., Gentner 1982). Even in languages where verbs occur in the prominent, sentence final position, or in isolation as the result of argument drop (e.g., Imai, Haryu, & Okada 2002), verb learning poses a serious challenge to children (but see Choi & Bowerman 1991; Choi 1998; Fenson, Dale, Reznick, Bates, Thal, & Pethick 1994; Tardif 1996; Tardif, Gelman, & Xu 1999). While there has been much speculation about what makes verb learning so difficult (see Golinkoff, Jacquet, Hirsh-Pasek, & Nandakumar 1996; Hirsh-Pasek & Golinkoff 2006 for a brief summary), the root of the problem remains unclear.

Gentner (1982; Gentner & Boroditsky 2001) was among the first to address why relational terms such as verbs are difficult to learn. She proposed that the difficulty might lie in the interplay between language and conceptual development. Under her 'natural partitions' hypothesis, nouns commonly refer to objects, and objects are naturally perceived as units; but verbs refer to relations between parts of events, and events can be conceptualized in terms of a variety of different relations.

With the help of the natural partitions hypothesis, we propose that the first problem in verb learning lies in the complexity of the interpretation of motion events. As compared to objects, actions are generally far more difficult to individuate. Verb learners are faced with the challenge of judging where one distinct action ends and another begins. Once the action has been individuated, the next task is to form a category of that action, and to recognize it across differing examples of actor, object, instrument, setting, etc. A child must learn to recognize that running is considered the same action whether it involves Grandma or Carl Lewis. Yet, these challenges are only the beginning; the process of mapping language to motion events introduces yet another level of complexity.

Actions are ephemeral, abstract, and fleeting (Gentner 1982; Gentner & Boroditsky 2001), and are often labeled before or after the target event has taken place (Tomasello & Kruger 1992). Since the verb label and the actual event do not necessarily occur simultaneously, children must infer the meaning of the novel word in order to attach it to the correct component of the motion event. Additionally, the very nature of actions stipulates that they necessarily occur in the presence of an entity (the actor or the object performing the action). Thus, children who hear action labels are faced with the problem of determining whether the label maps to the entity or to the action. The entity, if unnamed, often wins (Kersten & Smith 2002). Finally, verbs encode only part of what is happening in a motion event, including, but not limited to, *path* (the trajectory of an action with respect to some reference point, e.g., approach, enter), *manner* (how an action is carried out. e.g., walk, swagger, stroll), *result* (e.g., open, close), and *instrument* (e.g., hammer, shovel). Any event can involve a multitude of relations and it is the task of the verb learner to decide which relation in the event is the verb referent (Gentner 2003). Deciphering which aspect of the event is being labeled is often difficult, even for adults (Gillette, Glietman, Glietman, & Lederer 1999). Furthermore, there is much cross-linguistic variation in which elements of motion events are most likely to be verb referents (Talmy 1985). In the majority of languages, including Spanish, French, and Greek, *path* verbs are most frequent. For example, the most natural way to describe an event in which a woman ran as she exited a house would be to say the equivalent of *The woman **exited** the house* (*running*). However, in many other languages, including English, German, and Chinese, motion verbs more commonly express *manner*. In these languages it would be most natural to say the equivalent of *The woman **ran** out of the house.* Thus, learning motion verbs involves the disentangling of a variety of simultaneously occurring components of events, and deciding between a plethora of possible meanings (Gentner 1982; Golinkoff, Hirsh-Pasek, Mervis, Frawly, & Parillo 1995). When one considers the aggregate difficulty of interpreting motion events and subsequently mapping language to those events, it is no surprise that, in many languages, verbs appear later in children's vocabularies than nouns.

4.2 The cognitive linguistic approach to the problem of verb learning

Gentner's (1982; Gentner & Boroditsky 2001) natural partitions hypothesis provides us with a cognitive approach to verb learning. Extrapolating from Gentner's hypothesis, when children's prelinguistic concepts are aligned with the concepts expressed by words, learning those words would be naturally facilitated. There is much evidence supporting the idea that infants naturally perceive objects in ways that should help them learn nouns. All languages have words naming individual objects and categories of objects. Perhaps even at birth, infants perceive objects as units in their environments (Spelke 1990), and by 3 months can form at least some broader categories of objects such as animals versus vehicles (Arterberry & Bornstein 2001). But while research suggests that objects and object categories are present early in the infants' world view, there is no simple, universal conceptual organization for an event that could allow children to pick out all possible verb referents.

Research suggests that adults and children by the age of 7 possess flexible event concepts, with different languages favoring interpretations of motion events in terms of their manner or their path (Hohenstein 2001; Naigles & Terrazas 1998). Similarly, Choi and Bowerman (1991) report that from their earliest production of spatial terms, as early as 17 to 20 months, English- and Korean-speaking children extend spatial terms differently. Even at this early age, the way children talk about motion shows signs of language-specific patterns, reflecting differences in the concepts labeled by spatial terms in their languages.

Yet it is still an open question how and exactly when these event concepts come into place. One possibility, from the universal concepts perspective (e.g., Fodor 1975), is that children possess primitive, universal concepts of motion such that they naturally notice and attend to the elements of motion events that will enable them to learn verbs. Another possibility from the language-specific perspective (e.g., Whorf 1956) is that language focuses attention on certain aspects of the world so that children can form the necessary concepts to learn motion verbs in their language. Hybrid theories, on the other hand (e.g., Maguire, Hirsh-Pasek, & Golinkoff 2006), suggest that universal concepts initially help children focus on some particular aspect of an event, but that from early on children can use cues in their native language to limit the possible meanings of novel verbs. These cognitive linguistic hypotheses provide the foundation for our experimental endeavors. With the help of the infant methodologies discussed earlier, we are left with a series of testable questions: 1) Do young infants attend to the components of actions that words encode, such as *path* and *manner*? 2) If so, can infants categorize events on the basis of these concepts? 3) Finally, what happens when action meets word? How does language map onto these early event concepts? In the remainder of the chapter we will illustrate how habituation and the intermodal preferential looking paradigm have been used to provide evidence for the above questions.

4.3 Attention to potential verb referents in nonlinguistic events

The first experiment we will describe explores infants' processing of two of the elements of events most commonly encoded in motion verbs cross-linguistically – *path* and *manner*. While we know that objects are seen in ways that naturally promote noun learning (Spelke 1990), surprisingly little is known about how preverbal infants interpret motion events (but see Baldwin, Baird, Saylor, & Clark 2001; Casasola, Hohenstein, & Naigles 2003; Sharon & Wynn 1998). In order to examine this question experimentally, we need a methodology that is sensitive to infants' discrimination of subtly different motion events. Habituation is just that method. In this habituation experiment, we ask: In a *nonlinguistic* context, do preverbal infants notice path and manner in motion events (Pulverman 2005)?

Pulverman (2005) habituated English-learning 14- to 17-month-old infants to a motion event that included both a manner and a path. They were then tested with similar events in which the manner and/or path were systematically changed. Dishabituation to the changed events was measured to determine whether the infants had noticed the changes. Participants were assigned to either the High Vocabulary group or the Low Vocabulary group based on comprehension scores from the MacArthur Communicative Development Inventory (CDI) Infant Short Form (Fenson, Pethick, Renda, Cox, Dale, & Reznick 2000).

The habituation and test stimuli were computer-animated motion events featuring a purple starfish character performing an action and a stationary green ball on a black background. Each action included one of 3 manners (jumping jacks, spinning, or bending at the "waist") and one of 3 paths (over the ball, under the ball, or past the ball) (see Figure 2) resulting in 9 distinct actions (jumping jacks over, jumping jacks under, etc.). Crucially, no language accompanied these events.

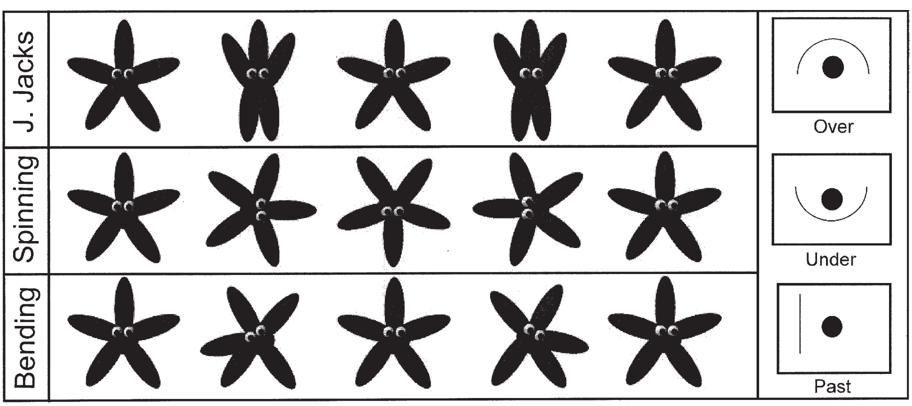


Figure 2. Examples of manners and paths used in stimuli. Although illustrated as a series of static postures, the starfish performed the manners as continuous motions.

In this experiment, participants were habituated to one of the nine stimulus events (e.g., jumping jacks over). Once habituated, each participant was presented with four test trials: a *control* trial with the same event as the habituation trials (e.g., jumping jacks over); a *path change* trial with the same manner as the habituation event but a different path (e.g., jumping jacks *under*); a *manner change* trial with the same path as the habituation event, but a different manner (e.g., *spinning* over); and a *both change* trial in which the manner and path were both different from those in all of the other events (e.g., *bending past*). A recovery trial of a laughing baby immediately followed the final test trial.

Results indicated that infants between 14 and 17 months old do in fact attend to and notice both manner and path changes, as evidenced by their dishabituation to an event in which only the manner or only the path, respectively, differed from the habituation event. Although noticing manner and path does not require conceptual knowledge of what manner and path are, in order for concepts of manner and path to be formed, these elements must be noticed. This evidence suggests that early in the verb learning process, infants are prepared to view the world in a way that will help them to learn verbs. This experiment additionally suggests that whether 14 to 17 month old infants specifically attend to manner is related to vocabulary level. Infants with higher vocabularies attended to manner, while infants with lower vocabularies in the same age group did not. Together with different findings from parallel research in Spanish, a language with a path verb bias, these results suggest that attention to event components influences lexical acquisition: Focusing in on the elements of motion events that are most relevant in one's language appears to promote vocabulary growth (Pulverman 2005).

Through this habituation experiment it has been shown that early in the verb learning process infants appear to be equipped with an essential tool for learning a wide variety of verbs. Results from a replication experiment indicate that even infants as young as 7 months of age, who have presumably learned very few words, are able to attend to and notice the elements of events encoded in motion verbs (Pulverman & Golinkoff 2003). Thus, with the help of the habituation paradigm it can be concluded that very early on, even before they learn their earliest words, infants are prepared to view the world in a way that will help them to learn not only nouns, but also verbs.

4.4 The categorization of path and manner in motion events

While the previous study provides us with information about infants' ability to attend to components of action, this is only the tip of the verb-learning problem. As Oakes and Rakison state, "words refer to categories of objects and events" (2003:4). That is, verbs label not *single* actions, but *categories* of actions and events. Previous research has demonstrated that infants can form at least some broad categories of objects at a very young age (Arterberry & Bornstein 2001). However, infants' ability to categorize motion events remains unclear. By using the preferential looking paradigm, the following study evaluates whether infants can form categories of path across multiple exemplars of manner, and categories of manner across multiple exemplars of path. This ability to form a category of both path, despite changes in manner, and manner, despite changes in path, is critical to the learning of relational terms. For example, to learn a preposition in English, infants must understand that "over" is the same path, irrespective of whether the person is *running* over, *jumping* over, or *crawling* over, etc. Similarly, in order to learn a verb in English, infants must recognize that "walking" is the same manner, irrespective of whether the person is walking *up*, walking *in,* or walking *under*, etc.

In order to test this categorization ability in infants, Pruden and colleagues (Pruden, Hirsh-Pasek, Maguire, & Meyer 2003) used a nonlinguistic variation of the preferential looking paradigm. The use of the preferential looking paradigm in this experiment is somewhat unique as compared to the methodology previously explained. In this experiment, researchers used a design that is not actually intermodal; here children were asked not to find the invariant across the visual and auditory modalities, but rather to display either a novelty or familiarity preference. This design is a modification of both the IPLP and the familiarization/novelty-preference procedure, a methodology often used in non-linguistic categorization tasks involving young infants (see Quinn & Eimas 1986).

Using this paradigm, Pruden et al. (2003) presented 7- to 9-, 10- to 12-, and 13- to 15-month-old infants with animated motion events in which the purple starfish from the previous study performed actions relative to a stationary green ball in the center of the screen. The actions were similar to those used by Pulverman (2005), each displaying a component of path and a component of manner. Infants were either exposed to four exemplars of exactly the same path across varying manners or four exemplars of exactly the same manner across varying paths. Importantly, no linguistic stimuli accompanied these events.

The experiment had four phases: 1) an introduction phase during which infants were introduced to the animated starfish moving on both sides of the split screen; 2) a salience phase during which any a priori preferences for the event clips to be used during the test phase could be measured; 3) a familiarization phase during which the infants were shown the four different event clips demonstrating an exemplar of the category being tested; and 4) a test phase to assess whether infants had formed the appropriate category. During test, infants were presented with two test events simultaneously on the split-screen. One side of the screen displayed a familiar test event, the other a novel test event. For example, when infants were asked to form a category of path across multiple examples of manner, the familiar test event involved the familiar path and a novel manner, while the novel test event involved a novel path and a novel manner. Throughout the experiment, infants' looking to each side of the screen was measured. Infants' ability to form a category for the desired event concept was identified by either a familiarity or novelty preference during the test events.

Results indicated that, by 10 months of age, infants can find the invariant path amidst differing manners, and by 15 months of age, infants can attend to an invariant manner amidst varying paths. Thus it appears that infants can isolate and abstract the invariant properties of path and manner even when other features of the display are changing. Further, a developmental trajectory appears such that path is detected prior to manner. Together with the previous habituation experiment, these studies suggest that infants detect nonlinguistic perceptual components of events (e.g., path and manner) that are codified in the world's languages. Additionally, infant show a nascent ability to categorize these components of motion events across changes in other components of motion events. Together, these studies with young infants show that, while verb learning is very difficult, at least some of the cognitive linguistic, conceptual foundations necessary for verb learning may be in place very early in life.

4.5 Use of universal and language-specific cues in verb learning

The previous experiments have helped us to identify the conceptual underpinnings of verb learning in young infants; however, it is unclear from these nonlinguistic tasks how the incoming language might affect infants' underlying concepts of motion events. In order to fully address verb learning from a cognitive linguistic perspective, language effects must be taken into consideration. Here, the IPLP becomes an extremely effective tool. For the next series of experiments, the question asked is: What happens to the universal concepts of path and manner when linguistic cues are added in a verb-learning task?

Maguire (2003) tested the interaction of universal concepts and language input at its earliest stages. Using the IPLP, 2- and 2 1/2-year-olds were presented with motion events involving an animated starfish again moving in relation to a stationary ball. On the basis of their comprehension of relational words (including verbs, prepositions, and adverbs) found on the MacArthur Communicative Development Inventory (Fenson et al. 1994), infants were divided into two groups – standard vocabulary and high vocabulary. In this series of IPLP studies, the experimental design followed exactly that which was described previously (Table 1). Infants were first introduced to the agent of the action, Starry. A salience trial followed in order to measure any a priori preferences for either test event. Infants were then shown the training phase. In this portion of the experiment, infants were given a novel label for Starry's actions. During each of the four training trials, Starry performed a novel action relative to the ball, accompanied by a novel label in the form of a sentence. *Initial*, *new label*, and *recovery* tests followed the training trials. Each experimental trial was separated by a centering stimulus to encourage the child to reattend to the screen and to center his or her gaze. The experimental stimuli were designed such that the novel label presented in training could refer either to the *manner* or to the *path* of the action. Which event component the infant believed the novel word to label was evidenced by the child's looking during the test portion of the experiment.

Experiment 1 asked the most basic question: Do 2- and 2 1/2-year-olds assume a novel label refers to the path or the manner of an action? Children heard sentences with a novel verb such as, "Look! Starry's blicking," as Starry repeatedly moved in a single manner across a single path. The results of Experiment 1 showed that high relational word vocabulary children looked significantly longer towards the target *path* during the initial test trials, indicating word learning and extension. However, the standard vocabulary children showed no preference. As there was no age difference between the children in each vocabulary group, this result suggests that the understanding of relational words, and *not age*, is a driving force in learning a novel verb label. As in previous experiments, a bias toward path was demonstrated. These results add support to the hypothesis that, because path is the more salient component in an action event, path might also be a comparatively more universal concept of action. However, the subsequent question arises, if children assume that every novel verb they hear is a path verb, how is it that they have manner verbs in

Table 1. The six experimental phases of the IPLP experiments of Maguire (2003)

Phase	Video	Audio	Display
Introduction	Stretch across	This is Starry! Meet my friend Starry. Starry is fun!	6 sec Repeated once.
Attention-getter	Baby's Face	Look up here!	3 sec
Salience*	Bend around Spin past	Look up here! What's Starry doing? What's going on up here?	6 sec
Attention-getter	Baby's Face	Hey blicking!	3 sec
Training	Spin around	Look! Starry's blicking! Do you see Starry blicking? Watch Starry blicking!	6 sec Repeated 4 times.
Attention-getter	Baby's Face	Find Starry blicking!	3 sec
Initial Test*	Bend around Spin past	Where's Starry blicking? Do you see Starry blicking? Look at Starry blicking!	6 sec Repeated 2 times.
Attention-getter	Baby's Face	Find Starry hirshing!	3 sec
New Label Test*	Bend around Spin past	Find Starry hirshing. Do you see Starry hirshing? Look at Starry hirshing!	6 sec
Attention-getter	Baby's Face	Find Starry blicking!	3 sec
Recovery Test*	Bend around Spin past	Blicking! Look at Starry blicking! Find Starry blicking!	6 sec

* Indicates visually identical stimuli

their productive vocabulary by 2 years of age (Fenson, et al. 1994)? Likewise, how are children learning anything in a language like English that is so dominated by manner verbs? To address these questions, Experiment 2 attempted to experimentally persuade toddlers to approach verb learning like English-speaking adults. Specifically, Experiment 2 asked, can 2- and 2 1/2-year-olds use multiple exemplars to guide verb learning?

The design of Experiment 2 was identical to that of Experiment 1 with the exception of one important factor: In Experiment 2 children saw a common manner across multiple exemplars of path. The linguistic stimuli were identical to those in Experiment 1. Experiment 2 thus asked whether a common label across multiple exemplars of a manner could shift children's focus such that they are now able to abstract and label the manner of the action as opposed to its path. This task was essentially that of Pruden et al. (2003) with the addition of a novel label given during the training phase.

The results of Experiment 2 showed that a common label for one manner across multiple paths does *not* provide enough information to shift children from labeling the path of the action to labeling the manner. In this case, neither high nor standard relational vocabulary children had a significant preference for labeling either path or manner. On the whole, children in Experiment 2 did not label either component of the action consistently. Thus, despite multiple exemplars, children were unable to interpret a novel verb as labeling the manner of an event. In response, Experiment 3 questioned whether *anything* could prompt children to approach a verb-learning task like English-speaking adults. Specifically, could multiple exemplars and additional syntactic cues help limit word meaning and thus guide verb learning?

The design of Experiment 3 was identical to that of Experiment 2 with the exception of the linguistic stimuli. In training, instead of hearing simply, "Look! Starry's blicking," children heard, "Look, Starry is blicking *over the ball*. Hey, now Starry is blicking *under the ball*, etc." Results of Experiment 3 suggest that when provided with both one manner across multiple examples of path, *and* syntactic cues, high relational vocabulary children mapped the novel word to the manner of the action. Standard relational vocabulary children, on the other hand, mapped the verb label to the path of the motion event.

Overall, it appears that 2- and 2 1/2-year-old children are unable to consistently label a single component of an action as the other changes. Follow-up studies, however have found that by 3 years of age, children finally adopt the English-speaking adult pattern of labeling the manner of a motion event. In order to ask these questions of older children, the IPLP task was adapted into a forced-choice experiment. Here, the audio portion of the experiment was read to the subject by the experimenter during the course of the video. Instead of utilizing visual fixation, children were asked to "Point to Starry blicking!"

Together, this series of IPLP studies suggests that children naturally map verbs onto the most salient universal action concept in a motion event. What helps English-speaking children move from a reliance on path to a language-specific reliance on manner are multiple exemplars and rich syntactic cues. These studies suggest that, although children are equipped with the ability to attend to and categorize universal action concepts in a motion event, crucially, they also make use of linguistic cues in their environment in order to hone in on verb meaning. Verb learning thus requires the coordination and integration of both *universal* and *language specific* cues.

5. Conclusion

Through the use of the habituation and intermodal preferential looking paradigms, progress is being made in understanding the process of verb learning. Based on results from these experimental paradigms, it appears that preverbal infants respond to *manner* and *path* in non-linguistic events. Furthermore, preliminary findings suggest that language-specific input has effects on these concepts during the course of language development. Uniting the universal concepts and language-specific perspectives of the relationship between language and cognition, our research has shown that both underlying concepts and language input are active and necessary in the task of verb learning. Research methods developed to ask questions of infants, combined with theoretical guidance from the cognitive linguistic perspective, suggest that children begin with foundational, universal concepts that are later integrated with language-specific input in the form of syntactic cues to influence the process of verb learning.

We began this paper by asking whether it is even possible to get inside the minds of preverbal infants. Our research demonstrates that collecting data on the conceptual underpinnings of language using preverbal infant participants is *not* a pipe dream. As we have shown, the habituation and intermodal preferential looking paradigms grant researchers access to the concepts from cognitive linguistics that are present in infants' minds and reveal how they are transformed as a result of cognitive development and increased exposure to a target language. We now have valid and reliable methodological tools that can be used to probe the very heart of the cognitive linguistic program in preverbal infants. The cognitive foundations of language, the effects of language input, and the coordination between the two can all be measured accurately and easily using the habituation and intermodal preferential looking paradigms.

References

Arterberry, M. E. & Bornstein, M. H. (2001). Three-month-old infants' categorization of animals and vehicles based on static and dynamic attributes. *Journal of Experimental Child Psychology, 80,* 333–346.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Biobehavioral development, perception, and action: Infants parse dynamic action. *Child Development, 72*(3), 708–717.

Barrera, M. E. & Maurer, D. (1981). The perception of facial expressions by the three-month-old. *Child Development, 52*, 203–206.

Bavin, E. L., Wales, R. J., & Kelly, H. (1998). Preferential looking: Testing structural knowledge. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research*, Vol 12. Stanford, CT: Ablex.

Bornstein, M. H. (1981). Psychological studies of color perception in human infants: Habituation, discrimination and categorization, recognition, and conceptualization. In L. P. Lipsitt (Ed.), *Advances in infancy research*, Vol. 1. Norwood, NJ: Ablex.

Bornstein, M. H. (1985). Habituation of attention as measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 253–299). Stanford, CT: Ablex Publishing Corp.

Bowerman, M. (1996). Learning how to structure space for language: A cross-linguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 385–436). Cambridged, MA: MIT Press.

Bowerman, M. & Choi, S. (2001). Shaping meanings for language: Universal and language specific in the acquisition ofr spatial semantic categories. In M. Bowerman & S. C. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge: Cambridge University Press.

Casasola, M. & Cohen, L. B. (2002). Infant categorization of containment, support, and tight-fit spatial relationships. *Developmental Science, 5,* 247–264.

Casasola, M., Hohenstein, J., & Naigles, L. R., (2003, April). Ten-month-old infants' discrimination of manner and path in motion events. In M. Casasola (Chair), *From infancy to adulthood: Exploring the effect of linguistic input on the discrimination of manner and path in motion events.* Symposium conducted at the Society for Research in Child Development Biennial Meeting, Tampa, FL.

Choi, S. & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition, 41,* 83–121.

Choi, S. (1998). Verbs in early lexical and syntactic development in Korean. *Linguistics, 36*(4), 755–780.

Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development, 14,* 241–268.

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2000). Habit2000: A new program for testing infant perception and cognition. (Version 1.0) [Computer Software]. Austin, the University of Texas.

Cornell, E. H. (1975). Infants' visual attention to pattern arrangement and orientation. *Child Development, 46*, 229–232.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science, 146,* 668–670.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs for the Society for Research in Child Development* (Serial No. 242).

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short form versions of the MacArthur Communicative Development Inventories. *Applied Psycholinguistics, 21,* 95–115.

Fodor, J. A. (1975). *The language of thought.* Cambridge, MA: Harvard University Press.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj, II (Ed.), *Language development Vol. 2: Language, thought, and culture* (pp. 301–334). Hillside, NJ: Lawrence Erlbaum Associates.

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind* (pp. 195–236). Cambridge, MA: The MIT Press.

Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In S. C. Levinson (Series Ed.) & M. Bowerman & S. C. Levinson (Vol. Eds.), *Language, culture, & cognition: Vol. 3. Language acquisition and conceptual development* (pp. 214–256). New York: Cambridge University Press.

Gillette, J., Glietman, H., Glietman, L., & Lederer, A. (1999). Human simulation of vocabulary learning. *Cognition, 73*, 135–176.

Golinkoff, R. M., Hirsh-Pasek, K., Cauley, K., & Gordon, L. (1987). The eyes have it: Lexical and syntactic comprehension in a new paradigm. *Journal of Child Language, 14,* 23–45.

Golinkoff, R. M., Hirsh-Pasek, K., Mervis, C. B., Frawly, W. B., & Parillo, M. (1995). Lexical principles can be extended to the acquisition of verbs. In M. Tomasello & W.E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 185–221). Hillsdale, NJ: Lawrence Erlbaum Associates.

Golinkoff, R. M., Hirsh-Pasek, K. & Schweisguth, M. (2001). The reappraisal of young children's knowledge of grammatical morphemes. In J. Weissenborn & B. Hoehle (Eds.), *Approaches to bootstrapping: Phonological, syntactic and neurophysiological aspects of early language acquisition*. Amsterdam and Philadelphia: John Benjamins.

Golinkoff, R. M., Jacquet, R. C., Hirsh-Pasek, K., & Nandakumar, R. (1996). Lexical principles may underlie the learning of verbs. *Child Development, 67*, 3101–3119.

Groves, P. M. & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review, 77,* 419–450.

Hirsh-Pasek, K. & Golinkoff, R. M. (1991). Language comprehension: A new look at some old themes. In N. Krasnegor, D. Rumbaugh, M. Studdert-Kennedy, & R. Schiefelbusch (Eds.), *Biological and behavioral aspects of language acquisition.* Hillsdale, NJ: Erlbaum.

Hirsh-Pasek, K. & Golinkoff, R. M. (1993). Skeletal supports for grammatical learning: What infants bring to the language learning task (pp. 105–124). In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research*, Vol. 8 (pp. 299–338). Norwood, NJ: Ablex.

Hirsh-Pasek, K. & Golinkoff, R. M. (1996a). The intermodal preferential looking paradigm reveals emergent language comprehension. In D. McDaniel, C. McKee, & H. Cairns (Eds.), *Methods for assessing children's syntax.* Cambridge, MA: MIT Press.

Hirsh-Pasek, K. & Golinkoff, R. M. (1996b). *The origins of grammar: Evidence from early language comprehension.* Cambridge, MA: The MIT Press.

Hirsh-Pasek, K. & Golinkoff, R. M. (Eds.). (2006). *Action meets word: How children learn verbs.* New York: Oxford University Press.

Hohenstein, J. M. (2001). Motion events similarities in English- and Spanish-speaking children. Unpublished dissertation, Yale University.

Imai, M., Haryu, E., & Okada, H. (2002). Is verb learning easier than noun learning for Japanese children? 3-year-old Japanese children's knowledge about object names and action names. *Proceeding of the 26th Annual Boston University Conference on Language Development,* 324–335.

Kersten, A. W. & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development, 73*(1), 93–109.

Lasky, R. E. & Spiro, D. (1980). The processing of tachistoscopically presented visual stimuli by five-month-old infants. *Child Development, 51,* 1292–1294.

Maguire, M. J. (2003). *Children's use of universal and language-specific cues in verb learning.* Unpublished doctoral dissertation, Temple University.

Maguire, M. J., Hirsh-Pasek, K., & Golinkoff, R. M. (2006). A unified theory of word learning: Putting verb acquisition in context. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 364–391). New York: Oxford University Press.

Mandler, J.M. (1991). Prelinguistic primitives. *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society,* 414–425.

Mandler, J. M. (2006). Actions organize the infant's world. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 111–133). New York: Oxford University Press.

McDonough, L., Choi, S., & Mandler, J. M. (2003). Understanding spatial relations: Flexible infants, lexical adults. *Cognitive Psychology, 46*, 229–259.

Naigles, L. (1990). Children use syntax to learn verb meanings. *Journal of Child Language, 17*, 357–374.

Naigles, L. R. & Terrazas, P. (1998). Motion-verb generalizations in English and Spanish: Influences of language and syntax. *Psychological Science, 9,* 363–369.

Oakes, L. M. & Rakison, D. H. (2003). Issues in the early development of concepts and categories: An introduction. In D. H. Rakison & L. M. Oakes (Eds.), *Early category and concept development: Making sense of the blooming, buzzing confusion* (pp. 3–23). New York: Oxford University Press.

Pruden, S. M., Hirsh-Pasek, K., Maguire, M. J., & Meyer, M. A. (2003). Foundations of verb learning: Infants form categories of path and manner in motion events. *Paper presented at the Boston University Annual Conference on Language Development*. Boston, MA.

Pulverman, R. (2005). *The relationship between language development and event processing: Lexical acquisition and attention to manner and path,* Unpublished doctoral dissertation, University of Delaware.

Pulverman, R. & Golinkoff, R. M. (2003). Seven-month-olds' attention to potential verb referents in nonlinguistic events. *Paper presented at the Boston University Annual Conference on Language Development*. Boston, MA.

Pulverman, R., Golinkoff, R. M., Hirsh-Pasek, K., & Sootsman Buresh, J. L. (under review). Manners matter: Infants' attention to manner and path in non-linguistic dynamic events.

Quinn, P. C., Adams, A., Kennedy, E., Shettler, L., & Wasnik, A. (2003). Development of an abstract category of representation for the spatial relation *between* in 6- to 10-month-old infants. *Developmental Psychology, 39,* 151–163.

Quinn, P. C. & Eimas, P. D. (1986). Pattern-line effects and units of visual processing in infants. *Infant Behavior and Development, 9,* 57–70.

Quinn, P. C., Cummins, M., Kase, J., Martin, E., & Weisman, S. (1996). Development of categorical representations for Above and Below spatial relations in 6- to 10-month-old infants. *Developmental Psychology, 39,* 151–163.

Ross, G. S. (1980). Categorization in 1- to 2-year-olds. *Developmental Psychology, 16,* 391–396.

Sharon, T. & Wynn, K. (1998). Individuation of action from continuous motion. *Psychological Science, 9*, 357–362.

Sokolov, Y. N. (1963). *Perception and the conditioned reflex* (S. W. Waydenfeld, Trans.). New York: Macmillian. (Original work published 1958)

Spelke, E. S. (1979). Perceiving bimodally specified events in infancy. *Developmental Psychology, 15*(6), 626–636.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14,* 29–56.

Spelke, E. S. & Hespos, S. J. (2002). Conceptual development in infancy: The case of containment. In N. L. Stein, P. J. Bauer, & M. Rabinowitch (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 223–246). Hillsdale, NJ: Erlbaum.

Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and the lexicon, Vol. III: Grammatical categories and the lexicon* (pp. 57–149). Cambridge, MA: Cambridge University Press.

Tardif, T. (1996). Nouns are not always learned before verbs: Evidence from Mandarin speakers' early vocabularies. *Developmental Psychology, 32,* 492–504.

Tardif, T., Gelman, S. A., & Xu, F. (1999). Putting the "noun bias" in context: A comparison of Mandarin and English. *Child Development, 70*(3), 620–635.

Tomasello, M. & Kruger, A. (1992). Join attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. *Journal of Child Language, 19*, 311–333.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf.* Cambridge, MA: MIT Press.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69,* 1–34.

Xu, F. & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74,* B1–B11.

Experimental study of first and second language morphological processing

Kira Gor

1. Introduction

Linguists are concerned with how language as a system works. Yet, some of them are especially interested in linguistic processing, or in other words, they wish to know what it is that people actually do when they produce and understand speech. And yet others set out to study linguistic processing in different populations of speakers – young children acquiring their first language (L1), child and adult second language (L2) acquirers, or children and adults with language impairments. The last two categories of researchers differentiate between a linguistic construct (rules postulated in a specific linguistic theory) and the actual mechanisms underlying speech production and perception, storage and retrieval of linguistic units, and conduct experiments to find out which theoretical constructs have a better fit to real speech processing observed in human speakers-hearers. In order to do so they utilize a variety of experimental techniques, which allow them to collect different types of speech data and perform analyses.[1] This paper will review several experimental paradigms used in research on the processing of inflectional morphology in different languages, and also different populations of speakers. It will discuss the experimental methods used in the studies, which contribute to the on-going debates concerning one of the most fundamental problems about human cognition, the problem of what kind of computations underlie linguistic processing and how the human mind deals with regularity and irregularity. Linguistic descriptions use symbolic rules, and formal linguistic theory assumes that linguistic processing, which is performed in the mind of the speaker-hearer, makes use of symbolic rules as well. At the same time, there is a whole direction

1. As many research paradigms dealing with human language, the study of linguistic processing is guided by two important considerations: how authentic and natural the collected speech data are, and therefore, how truly they represent the actual linguistic behavior, and which kinds of instrumental and statistical analyses can be performed on such data. And as in other domains, there is often a conflict between the two considerations, notably, the more authentic the data, say, a long segment of naturally produced spontaneous speech, the least likely it will lend itself to instrumental analysis and quantification. And the opposite is also often true; the constraints placed by some experimental procedures make the use of authentic speech virtually impossible.

in the study of language, which claims that symbolic rules are not part of the cognitive processes taking place in language production and perception, and these rules are only part of linguistic descriptions created by linguists. Instead of symbolic rules, human cognition relies on associative patterning in neural networks. According to this latter view, the linguistic mechanisms draw on the weights of connections between the nodes in a connectionist network, and not abstract symbolic rules.[2] The studies conducted in support of these two conflicting views make use of several research methods. This paper will describe (1) an elicitation technique used in a real and nonce verb generation task, and (2) a lexical decision task (LDT), which measures reaction times to different verb-forms.[3] Both these methods contributed significantly to promoting the understanding of morphological processing. The review of these methods will address specifically the role of different kinds of linguistic frequencies – type, token, whole-word, and stem-cluster – as well as phonological similarity, in morphological processing, and their use in linguistic experiments.[4] The paper will focus primarily on design and planning of experiments on the processing of inflectional morphology with a cross-linguistic perspective in mind and special emphasis on first and second language acquisition. Thus, it will provide the basic background knowledge necessary for future research into the mechanisms of morphological processing, while at the same time, paying attention to the theoretical underpinnings of such research and the impact of the obtained results on the current agenda in the debates on the status of symbolic rules in the processing of inflectional morphology. The experimental studies discussed below deal with three languages, which vary in the richness of inflectional paradigms, English, Italian, and Russian.

2. Two theoretical approaches to morphological processing

There are two main points of view with regard to morphological processing, which draw on two understandings of linguistic mechanisms in general. English past tense inflection has become the experimental ground for testing the predictions of the two conflicting points of view concerning morphological processing with most of the data, especially, generated in connectionist modeling, coming from English. Past-tense verbs in English are inflected in two different ways. The vast majority of verbs forms their past tense by a general-purpose concatenative rule, 'add *-ed* to the stem,' which constitutes regular inflec-

2. It is also important to mention that the nativist position strongly supported by formal linguistics is based on the idea of innateness of language. Connectionist modeling places emphasis on learning and studies the impact of different kinds of input on language learning.

3. Nonce words are created by experimenters by manipulating the properties of real words; they are novel to the native speaker, do not have any meaning, but may be assigned a fictitious meaning within the experiment.

4. See the section Input Frequencies and Probabilities in Linguistic Processing below for the definitions of different kinds of frequencies.

tion: *walk-walked, play-played*.[5] A relatively small group of irregular verbs is inflected by some unpredictable stem change with or without the addition of some kind of inflection.[6] These irregular verbs can cluster into 'neighborhoods' with similar stem allomorphy, such as *sing-sang, ring-rang, spring-sprang.*[7] These properties of English, namely a categorical distinction between regular and irregular past-tense inflection, make it possible to investigate the issue whether regular and irregular processing are performed by two distinct mechanisms or just one, and whether regular, but not irregular inflection makes use of symbolic rule computation.

According to the dual-system approach (Clahsen 1999; Jaeger et al. 1996; Marcus et al. 1992, 1995; Pinker 1991, 1999; Pinker & Prince 1988, 1991, 1994; Pinker & Ullman 2002; Prasada & Pinker 1993; Ullman 1999, 2001), which is most consistently represented in Steven Pinker's Words and Rules Theory, linguistic processing is subserved by two main systems – the mental lexicon and a computational system (Pinker 1999). Words in the mental lexicon are connected to each other based, e.g., on phonological or semantic similarity, and their storage and retrieval are performed in associative memory; symbolic rules are applied by computation. In the domain of inflectional morphology, irregular forms are stored in the mental lexicon while regular forms are computed on-line. There is no need to store regular inflected forms, as they will be assembled by the application of concatenative rules. However, no such rules can exist for the irregular forms, therefore they have to be memorized, and retrieved from the mental lexicon when needed. For example, English speakers do not need to store both the singular form of the noun *tree* and its plural form *trees,* since the plural is generated by the application of a regular rule: add *-s* to the singular stem. The noun *child*, however, needs to be memorized together with its irregular plural *children*. One important implication arising from such a standpoint is the assumption that irregular forms processed in associative memory will show frequency effects. The nature of these frequency effects will be discussed more in detail in the next section. Generally speaking, the underlying assumption in research on frequency effects is that the words that are activated more often have stronger memory traces, and as a result, are retrieved from memory faster. Contrary to irregular inflection, regular inflection will not show any frequency effects, because symbolic rules are applied when the necessary criteria are met, regardless of the frequency of use. Since regularly inflected words are assembled on-line, they are not stored in undecomposed form (with inflections attached to the stems), and thus cannot show whole-word frequency effects. Thus, the dual-system approach makes clear predictions about the connection between input frequencies to the speaker (learner): frequency effects will be observable in irregular inflection, but not in regular inflection. One of such predicted effects includes shorter reaction times (RTs) in

5. Concatenation is addition of linguistic elements to one another, "stringing" them in a bead-like fashion.

6. Note, however, that some past-tense verb forms are identical to the basic stem, e.g., *put-put.* Such no change verbs are also irregular because they do not have the *-ed* inflection.

7. Allomorphy refers to variations in the phonological composition of morphemes; these phonologically different variants of the morpheme are called allomorphs.

a lexical decision task (LDT) to high-frequency than to low-frequency irregular word-forms, but no such effect for regular ones. The same strong predictions involve the role of phonological similarity in regular and irregular inflection – only irregular word-forms will show effects of phonological similarity, e.g., in a nonce word generation task. Speakers will use the irregular inflectional pattern only for the nonce words, which are very similar to the existing irregularly inflected words, whereas they will use the regular inflectional pattern regardless of phonological similarity. Over time, the dual-system approach moved from its strong version, according to which no regular inflected words are stored in the lexicon in undecomposed form, to a weaker version expressed in the Words and Rules Theory (Pinker 1999). The Words and Rules Theory admits that a certain number of high-frequency regular word-forms can be stored undecomposed, but that this fact does not affect the general processing model.

The single-system approach is built on the premises that there is no categorical distinction between two different processing mechanisms, the rule-based and the association-based ones (Bybee 1985, 1995, 2002; Langacker 1987, 1988; MacWhinney & Leinbach 1991; Plunkett & Marchman 1991, 1993; Rumelhart & McClelland 1986). All the words, uninflected, regularly and irregularly inflected, are stored in the mental lexicon and form associations based on phonological and semantic similarity. The frequency of these mappings between word-forms in the lexicon is crucial for their processing, e.g., for the speed of their retrieval. If a certain type of mapping, e.g., between the verb stem and the regularly inflected past-tense form, *repeat–repeated, discuss–discussed*, etc., is highly frequent in the network, such pattern of association gains in strength. This associative patterning has been modeled in numerous versions of connectionist networks, in which the weights of connections between the nodes representing word-forms are adjusted depending on the frequency of their activation. Obviously, the single-system approach predicts that both regular and irregular inflection will show frequency effects, since both regular and irregular word-forms are processed in the same associative network. It goes without saying, that phonological similarity lies at the core of single-system models, and its effects are predicted both for regular and irregular inflection. Associative patterning is a proposed alternative to symbolic rule computation, and the single-system approach seeks to show that our mind processes language without symbolic rules. Several research directions, most of them involved in connectionist modeling, espouse the single-system approach, including emergentism, a relatively recent direction, focusing on the emergence of language in an associative network in children acquiring their L1 as a product of input processing (Elman et al. 1996).

Since English does not possess a developed conjugational paradigm, and has only one large regular productive default class and one small irregular unproductive class in the past tense, in other words, it is impossible to divorce regularity, productivity, and high type frequency in it, research on other languages with complex inflectional morphology is in order. First of all, let us briefly define the terms regularity, productivity, type frequency, and default. Regularity in inflectional morphology refers to a pattern involving less changes to the shape of the morphemes involved. Regular inflectional processes can be described by a simple concatenative rule: add a certain ending to the stem. English past-tense and noun plural formation are examples of regular inflectional processes.[8] Productivity refers to the fact that the pattern is used in new words, e.g., English uses the productive *-ed* pattern in *blog-blogged*. Type frequency refers to the frequency of occurrence of a linguistic pattern, or in other words, to the size of a certain class of words using this pattern. And finally, default is characterized by the most open schema, and it is used "when all else fails" (Bybee 1995: 452). It is important to maintain this distinction in the data, because the dual- and single-system theories connect regularity to different linguistic parameters. Thus, the dual-system approach claims that regularity and productivity are interconnected, and therefore, regular inflection has to be productive. Also, it does not explicitly differentiate between regular and default processing, often using the 'regular' and 'default' terms interchangeably. Contrary to that, the single-system approach draws a connection between productivity and high type frequency, and maintains that regularity, productivity, and default are all different parameters.

The cross-linguistic data on the processing of verbal morphology in German, Italian, Russian, Norwegian, and Icelandic, languages with a complex conjugational paradigm and often more than one regular verb class, by adults, children, as well as L2 learners, has shed some new light on the disputes (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005; Clahsen 1999; Matcovich 1998; Orsolini & Marslen-Wilson 1997; Orsolini, Fanari, & Bowles 1998; Ragnasdóttir, Simonsen, & Plunkett 1997; Simonsen 2000). All these studies have found the influence of frequency on the processing of all the verb classes, including the regular default classes, both in children and adults. It should be noted that a series of studies on the processing of German inflectional morphology supports the opposite, dual-system model of morphological processing and demonstrates the differences in the processing of different classes of German participles and noun plurals that are consistent across different experimental paradigms (Clahsen 1999). However, these studies do not maintain the distinction between regular and default inflection, which strongly undermines their claims. Research on the processing of Russian verbal morphology by native Russian adults and children, and adult American speakers of Russian as L2, which will be discussed more in detail below, has produced several results incompatible with the categorical distinction between regular and irregular morphological processing (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005). Thus, in nonce verb generation, all the groups of subjects generalized several conjugational patterns ranging in regularity; type frequency was one of the determining factors in the choice of the conjugational pattern – high type frequency patterns were generalized to novel verbs more readily; in a LDT, RTs were shorter for high-frequency forms in the conjugational paradigm. Generally speaking, the data on verbal processing in several languages other than English indicate that these languages have more than one generalizable regular inflectional pattern or several patterns differing in the degree of regularity, and do not support a categorical distinction between regular and irregular morphological processing.

---

8. Of course, the *-ed* inflection is pronounced differently depending on the final phoneme of the stem, compare *stop-stopped, rob-robbed, and visit-visited.*

3. Input frequencies and probabilities in linguistic processing

This section will briefly review the kinds of linguistic frequencies discussed in the literature, and the assumptions about the connection between frequencies and probabilities and their role in linguistic processing. This will prepare us to follow the experimental design of the studies using different experimental methods reviewed below. *Token frequency* is the frequency of the word-form in speech. In other words, token frequency tells us how often a language user encounters a certain word, either by producing it or hearing it produced by other speakers. E.g., all the instances when a given individual has heard the word *spoon* constitute token frequency of the word *spoon* in the input to this individual. Obviously, the real frequencies, either for an individual speaker or for a population of speakers, are unavailable; we can only rely on estimates, which are based on language corpora. At present, few large representative corpora are available for languages other than English. The largest and most used corpus of English, COBUILD, is constantly growing; in 2004 it amounted to 524 million words.[9] Another widely used source is the CELEX database, which in fact, comprises three different searchable databases, the Dutch, English, and German ones.[10] One of the general challenges for the creators of such databases and language researchers alike is relatively limited availability of transcribed spontaneous speech data, and the necessity to rely extensively on written texts. This certainly limits the reliability of corpus-based estimates. Usually, the standard index for word frequency is the frequency of occurrence per 1 million word usages.

The rationale behind using token frequency counts in psycholinguistic experiments is that the more often the word stored in the lexicon is used, the stronger the memory traces for this word. This will result in greater access speed for this word. For simple undecomposable words consisting of a bare stem and no inflection token frequency is uncontroversial. However, the question arises concerning the psycholinguistic mechanisms underlying storage and access of complex inflected word-forms. It is hypothesized that *whole-word frequency,* or the frequency of the inflected word-form, should matter only for the complex words, which are stored undecomposed. This, in turn, implies that whole-word frequency effects (e.g., shorter RTs to high-frequency word-forms in a LDT) signal that the inflected word is stored in the mental lexicon, and not computed on-line by a symbolic rule. Then, *stem-cluster frequency*, or the combined frequency of all the word-forms with the same stem should reflect the role of the stem.[11] Its effects should signal that the word is stored decomposed, and inflected by symbolic rule computation. To give an English example, whole-word frequency for the inflected verb *walks* is a fraction of the cumulative stem-cluster frequency for this verb, which includes the forms *walk, walks,*

---

9. Information on COBUILD can be found at the website at http://www.cobuild.collins.co.uk.

10. The CELEX database is maintained by the Dutch Center for Lexical Information at the Max Planck Institute for Psycholinguistics, Nijmegen. The reader is referred to the website at http://www.ru.nl/celex/ for further information.

11. Whole-word frequency is sometimes referred to as surface frequency and stem-cluster frequency as lemma frequency.

walked, and *walking*. A comparison of stem-cluster and whole-word frequencies was used in the study of English past-tense inflection (e.g., Alegre & Gordon 1999).

Type frequency refers to the frequency of occurrence, or in other words, size of a certain class of words in a language, and it is established based on dictionary counts. For example, one can estimate the type frequency of masculine and feminine nouns in French. Importantly, all the words in a class share some common feature(s), e.g., Russian 1st conjugation verbs use the thematic vowel 'e' in their non-past paradigm, while 2nd conjugation verbs use the vowel 'i'. Therefore, type frequency can also refer to a pattern or rule, depending on the theoretical position. This ultimately means that the effects of type frequency in linguistic processing do not support the dual- or single-system approach by themselves unless we can independently demonstrate that the effects refer to a symbolic rule, or alternatively, to an associative pattern.

And finally, *frequency of use* has had limited application in research on the processing of inflectional morphology. This measure is 'intermediate' between token and type frequencies, as it refers to all the uses ('tokens') of a certain class of words in speech. This measure can be obtained by recording all the instances of the use of a certain class of words in an observational experiment.[12] A similar measure, labeled *the number of uses*, made it possible to obtain a more realistic assessment of the input to L2 learners of Russian (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2005). In a population of speakers with limited input and incomplete access to the linguistic system, such as children acquiring L1 or L2 and adult L2 learners, this combined measure is potentially useful. The number of uses was also a basis for the assessment of the rule weight in the Rule Competition Model of child L1 acquisition of English past-tense morphology (Yang 2002).

The role of frequencies in linguistic processing is not restricted to stronger memory traces for high-frequency word-forms, which increase the speed of access to linguistic units stored in the lexicon and their retrieval. Note that this understanding focuses on the role of token frequency; however, the hypothesized role of type frequency is based on somewhat different assumptions. Thus, the strength of association between two word-forms in the lexicon depends on the frequency of activation of the connection between them. It is believed that human speakers store the information about different kinds of linguistic frequencies they encounter. The information about both the frequency of linguistic events and the frequency of combinations of events is hypothesized to be built into the probabilistic mechanisms used both in speech production and reception (Bod et al. 2003). The influence of probabilities associated with type frequency manifests itself in nonce verb generation, when the most frequent pattern has the highest generalization rates (Chernigovskaya & Gor 2000). Thus, input frequencies and probabilities based on them influence linguistic processing and are psycholinguistically interconnected.

The idea that rules have probabilities in young children acquiring their native language, whose linguistic system has not yet stabilized and who operate with several competing grammars, was explored by Charles Yang (2002). Yang proposed the Rule Competition Model of child L1 acquisition of English regular and irregular verbal morphology, which is

12. The study of the American learners of Russian discussed below relied on an estimate based on a simulation using the instructional materials (Chernigovskaya & Gor 2000).

a radical departure both from the dual- and single-system approaches. The new branch of linguistics, Probabilistic Linguistics, explicitly explores the idea that the language faculty is inherently probabilistic and studies frequency effects (Bod et al. 2003). Based on the results of several experiments on the processing of Russian verbal morphology by child and adult L1 speakers as well as adult American learners of Russian (Chernigovskaya & Gor 2000; Gor, 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005), the Rules and Probabilities Model for both native and non-native processing of complex Russian verbal morphology was proposed (Gor 2003, 2004). This model is similar to Yang's Rule Competition Model when it addresses L2 morphological processing, however, the idea that rules have probabilities in native adult processing of Russian verbs, both in production and perception, takes the Rule Competition Model one step further. In conclusion, several research agendas point in the same direction, notably, that the categorical distinctions between symbolic rules and the lexicon, between regular and irregular morphological processing, and the strong claim that in linguistic processing, symbolic rules are immune to probabilities need reassessment in the light of new data.

4. Real and nonce verb generation task

Research on several languages, including English, Italian, Russian, German, Norwegian, and Icelandic, with adults, children, and L2 learners, uses the same experimental paradigm, a verb generation task, to study the processing of verbal morphology (Chernigovskaya & Gor 2000; Gor 2003, 2004; Gor & Chernigovskaya 2001, 2003a, 2003b, 2005; Clahsen 1999; Matcovich 1998; Orsolini & Marslen-Wilson 1997; Orsolini, Fanari, & Bowles 1998; Prasada & Pinker 1993; Ragnasdóttir, Simonsen, & Plunkett 1997; Simonsen 2000). The experiments adapt an elicitation technique first developed in a study by Bybee and Slobin (1982): The experimenter works individually with each subject and elicits the expected verb-form by acting out a short quasi-dialog. In some cases, instead of oral presentation, adult subjects receive a written list of sentences and are asked to fill in the blank with the appropriate form of the provided verb in writing (Orsolini & Marslen-Wilson 1997). For obvious reasons, with native adults only nonce verbs are used for the testing material, as with real verbs one would expect ceiling effects in this population of speakers, while with children and L2 speakers, whose internalized verbal system is incomplete, both existing and nonce verbs are used. The aim of such experiments is to study the factors, which determine the generalization rates, or the rates of responses using a particular conjugational pattern. The list of such possible factors includes type frequency, regularity, default, and phonological similarity to the existing verbs, such as the rhyming effect. The following sections will discuss the methodology and the main findings of the studies involving two languages, Italian and Russian.

4.1 Nonce verb generation task with adult L1 speakers of Italian: The role of phonological similarity

The experiment on Italian nonce verb generation explored the effect of two variables on the relative generalizability of inflectional processes: (1) regular versus irregular conjugation class membership, and (2) phonological similarity to the existing Italian verbs (Orsolini & Marslen-Wilson 1997: 22).[13] This study with adult native speakers of Italian as subjects was designed similarly to an earlier study on English past-tense inflection (Prasada & Pinker 1993), and replicated its task conditions, since the goal was to verify if the results obtained in the English study held for Italian. The original study by Prasada and Pinker was done in the framework of the dual-system approach and tested the hypothesis that in nonce verb generation, the regular *-ed* rule would be applied regardless of the similarity of the nonce verb to the existing regular verbs. Unlike the regular rule, irregular conjugational patterns will be applied only to the nonce verbs bearing a close resemblance to the existing irregular verbs. And indeed, the obtained results confirmed the hypothesized predictions. Both the English and Italian study explored the influence of phonological similarity on regular and irregular nonce verb generation, and given that in Italian the differences between irregular and regular past-tense inflection are smaller than in English, a less categorical disjunction between regular and irregular inflection was expected for Italian past-tense verbs. The choice of the stimulus material reflects the goals of the study. The nonce verbs created for the experiment belonged to irregular conjugation II (infinitive ending in *-ere*) and regular conjugation III (infinitive ending in *-ire*) classes (Orsolini & Marslen-Wilson 1997: 22ff.). Three phonological similarity conditions were defined to match the conditions of the Prasada and Pinker study. In Condition 1 (high similarity), the verbs had minor phonological changes in the initial syllable and closely resembled the existing verbs. For example, the nonce verb *frendere* is very similar to the existing irregular verbs *prendere* and *rendere.* In Condition 2 (root similarity) the nonce verbs did not match closely any of the real Italian verbs, however, the initial part of the verb as well as the final root segment resembled the existing Italian verbs. For example, the initial syllable *imm-*, which is used by verbs like *immigr-are* and *immett-ere*, was used for the nonce irregular verb *immund-ere*. This item has the highly frequent *-nd* root ending typical of irregular conjugation II verbs. And finally, in Condition 3 (low similarity) the nonce verbs did not use the initial or final segments of any existing Italian verbs, an example of such a verb is *maffecere*. The likelihood that a certain conjugational pattern is chosen may depend not only on how close the created pseudo-verb is to the existing verbs, but also on how many similar-sounding verbs there are in the language. Thus, to match the study of English past-tense generation (Prasada & Pinker 1993), the "rhyming rate" of pseudo-regular and pseudo-irregular verbs was measured for each similarity condition and comparable rhyming rates were obtained for most Italian stimuli sets. The rhyming rate was defined as the average number of existing verbs, which rhymed

13. In addition to the nonce verb elicitation task discussed here, the article reports the results of a cross-modal immediate repetition priming task experiment. For lack of space, they will remain outside the scope of this paper. The reader is referred to the original article for details (Orsolini & Marslen-Wilson 1997).

with the novel verbs. The way phonological similarity and the size of the neighborhood (similarly-sounding verbs) are calculated depends on the language in question. Note that, generally speaking, the strength of a neighborhood may be represented as a figure obtained by subtraction of the number of enemies from the number of "friendly" neighbors (see Ullman 1999:56). The enemies are similarly-sounding verbs with a different conjugational pattern, or in case of English past-tense inflection, with a different stem-past mapping, e.g., for the stem-past pair *sing-sang, ring-rang* would be a friendly neighbor, while *bring-brought* would be an enemy. This type of computation was used by Michael Ullman in an experiment, which studied the influence of phonological similarity on acceptability ratings of English regular and irregular stems and past-tense forms by adult native speakers (Ullman 1999).

An equal number of pseudo-irregular (ending in *-ere*) and pseudo-regular (ending in *-ire*) verbs were created, 20 in each of the three phonological similarity conditions, thus the total number of experimental stimuli was 120. The stimuli were presented as infinitives with fictitious definitions, e.g., the nonce verb *effadere* was defined as "to cook nervously." This study used the written format for data collection, in which the subjects had to fill in the blanks in the sentences by generating the past participle of the provided verb. Here is an example of such a sentence, where the verb *effadere* was to be used: *The fish soup was not very tasty because Maria had* ________ *and had forgotten to put salt in it.*

20 adult native speakers of Italian, paid volunteers, took part in the experiment. The subjects' responses were further categorized as regularizations (i.e., the past participles produced using regular conjugation III and II[14] thematic vowels and suffixes), hyper-regularizations (the forms with a default conjugation I vowel and suffix in response to conjugation II and III verbal stimuli), and irregularizations (the forms with stem changes and irregular suffixes). Based on the obtained results, the study arrived at the conclusion that, indeed, the sharp dichotomy between regular and irregular mechanisms of morphological processing postulated by the dual-system theory cannot handle the data on Italian verb generation. First, the rate of irregularizations, or the use of the irregular conjugation II pattern was much higher for Italian than for English. Irregular conjugational patterns were generalized at a high rate even in the root similarity condition, where there was no close similarity to the prototype. And second, regular conjugation III inflection in subjects' responses exhibited the influence of phonological similarity, which should not have been present in regular verb generation. The study failed to support the single-system account either, thus leaving the answer about the type of processing observed in Italian verb generation open. One important contribution of this study to the dual- versus single-system debates and our understanding of the mechanisms of morphological processing is that it demonstrates that the study of one single language is insufficient to support any universal claim about linguistic processing, and that cross-linguistic data are in order.

14. It should be noted, that not all conjugation II verbs are irregular in Italian.

4.2 Real and nonce verb generation task with adult L1 and L2 speakers of Russian and L1 children: The role of type frequency[15]

While the Italian study discussed above dealt with the role of phonological similarity, the experiments on Russian verb generation focused on the interaction of regularity and type frequency, or class size. They addressed the opposite claims about the role of frequency in regular versus irregular inflection put forward by the proponents of the dual- and single-system approaches. According to the dual-system account, regular inflection, unlike irregular inflection, is unaffected by input frequencies. The single-system account maintains that both regular and irregular inflection are influenced by input frequencies. A series of experiments on Russian verb generation has prompted the Rules and Probabilities Model (Gor 2003, 2004), which challenges both the dual- and single-system approaches. The model is based on the following claims: (1) In languages with numerous inflectional patterns regularity can become a gradual parameter, and as a result, a categorical distinction between regular and irregular inflection is unjustified. (2) In the processing of regular inflectional morphology, speakers make use of abstract rules. (3) Rule application is influenced by the probabilities based on linguistic frequencies. In other words, the data on several populations of L1 and L2 speakers of Russian suggest that regular inflection is influenced by input frequencies, but it is rule-based rather than associative memory-based. The following sections will describe the experiments, which demonstrated the role of input frequencies and probabilities in regular morphological processing. Another series of experiments, not discussed here (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2001), showed that regular morphological processing in Russian relies on rule application rather than associative patterning (Gor 2004). The next section will provide the basic information about Russian verb conjugation and different kinds of frequencies used in designing the experiments with native and non-native speakers of Russian. It will exemplify the kind of issues the researcher deals with when working with material on an individual language.

4.2.1 *Russian verbal morphology and the resources on Russian verb frequencies*

The debates on the processing of English past-tense morphology early on identified a major 'weakness' of the English data, notably, the impossibility to separate regularity, productivity, high type frequency, and the default. Indeed, 96% of English verbs are regular and productive, while only 4% are irregular and unproductive. Another feature of the English past tense is that it has only one regular conjugation rule, and therefore regular and default processing are merged together. These limitations led researchers to look at other languages with rich inflectional morphology. As it turns out, Russian has strong advantages over English in that it has numerous conjugation classes differing in the degree of regularity and size. The on-going collaborative project on the processing of Russian verbal morphology by different groups of L1 and L2 speakers, children and adults, as well as children and adults with language disabilities, conducted at the University of Mary-

15. For a more detailed account of the verb generation task with L1 children and L2 adults see Gor and Chernigovskaya, 2003a, b, 2005.

land, USA and Saint Petersburg State University, Russia capitalized on these features of the Russian conjugational system. This section will discuss the major features of the Russian conjugational system necessary in order to follow the experimental design and evaluate the proposed argumentation. It will also review the resources on verb frequencies used in the study.

According to the one-stem verb system created by Roman Jakobson (Jakobson 1948), Russian has 11 verb classes, each with its own suffix, or verbal classifier. The eleventh class has a zero suffix, and is subdivided into smaller subclasses depending on the quality of the root-final consonant. This is a small class, especially given the variety of conjugational patterns it includes, and there are well under 100 basic stems in it (Townsend 1975). The conjugational patterns of some of the sub-classes of the non-suffixed stems have idiosyncratic features, and thus form verb clusters, which can be compared to the neighborhoods of English irregular verbs, or alternatively, characterized by the minor rules. The remaining 10 suffixed classes are identified by the suffix: -aj-, -ej-, -a-, -e-, -i-, -o-, -ova-, -avaj-, -nu-, and -zha-.[16] The suffix determines all the parameters of the conjugational paradigm, including the choice of the thematic vowel in the inflections (*-e-* for 1st and *-i-* for 2nd conjugation), and different types of stem changes (see Table 1). When the endings are added to the stem (which includes the optional prefix, the root, and the suffix), an automatic truncation rule works at the juncture of the stem and the ending. If the stem ends in a vowel and the ending begins in a vowel, the first vowel is truncated. The same is true for the consonants: the first one is deleted. Past tense and infinitive endings begin with a consonant, and non-past tense endings begin with a vowel, therefore, stem-final vowels will be deleted in non-past tense forms, and consonants will be deleted in past tense and infinitive forms. The examples below illustrate the automatic truncation processes taking place when the inflections are added to the stems in two verb classes, the -aj- and -a- classes, used in the experiments.

Table 1. Non-automatic morphological processes in the stems included in the experiments

Verb classes and sub-classes	-aj-	-a-	-ej-	-e-	(i)j-*	-i-	-ova-	-avaj-	(o)j-
Consonant Mutation		√		√		√			
Stress shift[17]		√		√		√			
Suffix alternation							√	√	
Vowel alternation									√
Vowel deletion					√				

* The (i)j- and (o)j- stems are the sub-classes of zero-suffixed stems, each of them has an idiosyncratic ("irregular") feature in the conjugational pattern.

16. The consonant "zh" represents any palatal consonant – a hushing or "j" – and is not part of the suffix.

17. This paper does not discuss stress shifts in the obtained data. Unlike all the other morphological processes that can be predicted given the verb stem, stress shifts need to be lexically encoded. However, the pattern of stress shift within the paradigm is fixed.

gul'-aj- + t' = *gul'at'* (infinitive, "walk")
gul'-aj- +u = *gul'aju* (1st person sg, non-past tense, "I walk")

pis-a- + t' = *pisat'* (infinitive, "write")
pis-a- + u = *pishu* (1st person sg, non-past tense, "I write")

Note that the infinitive does not contain the information about the verb class, because the *j* of the suffix may have been truncated and is unrecoverable. This means, that a novel infinitive ending in *-at'* may be interpreted either as an -aj- or -a- verb. This property of the Russian verbal system makes it possible to research the role of probabilities (and phonological similarity as well) in verb generation. Also, note that in the -a- stem non-automatic consonant mutation takes place in addition to automatic truncation.

Overall, the Russian verbal system possesses the following features:

– Numerous verb classes varying in size;
– Developed conjugational paradigm;
– No sharp division between regular and irregular classes;
– Several regular classes in addition to default;
– Infinitives of many verb classes have unrecoverable stems due to the truncation of the stem-final consonant before consonantal endings.

Table 1 lists the morphological processes shaping the conjugational patterns of the stems chosen for the experiments. It does not include automatic consonant or vowel truncation, which occurs at the juncture of the stem and the ending.

Table 2 contains information about the type frequencies of the stems included in the experiment and about their productivity. Since no such data were available, we performed the computations ourselves on the *Grammatical Dictionary of the Russian Language* (Zalizniak 1980) with approximately 100,000 entries. In the first row of Table 2 corresponding to the Russian language, the first set of numbers in each column represents the results of our verb counts. These counts contain all the verbs belonging to a particular conjugation class and include the prefixed verbs.[18] For the small unproductive classes, we also provide the number of unprefixed basic stems below (based on Townsend 1975; and Davidson et al. 1996). The second and third rows contain two kinds of data on input frequencies specifically obtained for L2 learners taking part in the verb generation experiment. If we assume that the type frequencies found in the Russian language are available to adult native speakers through their linguistic experience, the situation with L2 learners is much more complicated, since it is impossible to exactly estimate the input they have received during their exposure to L2. Therefore, we decided to focus on the L2 learners, for whom we could produce the most accurate estimates of the input frequencies, the American students of Russian who had completed one year of study of Russian at the University of Maryland. We simulated all the language activities in which they were engaged, in class and at home, based on the set of instructional materials they were using. Thus, the estimates assume that the learners completed all the assignments, which was certainly not

18. Natalia Sliusar from the Department of General Linguistics at St. Petersburg State University performed these computations.

Table 2. Type frequencies of the verb classes included in the experiments: Native and second language input[19]

Verb classes	-aj- Productive	-a-	-ej- Productive	-e-	(i)j-	-i- Productive	-ova- Productive	-avaj-	(o)j-
Russian language Type frequency	11814	940 Appr. 60 stems	608	328 Appr. 50 stems	160 7 stems	7019	2816	94 3 stems	98 5 stems
Input to L2 learners Type frequency	55 (86)[20]	14 (24)	0 (4)	8 (12)	3 (3)	52 (80)	13 (34)	2 (7)	2 (5)
Input to L2 learners Number of uses	4333	1298	12	782	239	4546	555	273	158

exactly true, however, it was a good approximation of their linguistic input, because the curriculum expects them to follow the textbook very closely. Naturally, in more advanced speakers, who have more exposure to the L2 outside the classroom and the textbooks, such estimates become less accurate.

At the time of the experiments, there existed two searchable internet-based Russian corpora, the Uppsala Corpus and the Tübingen Corpus, both available for on-line query at the site of the University of Tübingen.[21] The Uppsala Corpus contains 1 million words and comprises literary texts, press, and scientific texts from the second half of the 20th century. The Tübingen Corpus contains only 600,000 words, however, it is very valuable, since it contains only the oral interview data (presumably edited) – these are the interviews published in Russian press of the end of the 20th century. According to the estimates of Lönngren, a corpus of 1 million words is reasonably representative for the words with the frequency of 80 per million and higher (Lönngren 1993), but not for lower frequency ones. The list of such words is published in the *Frequency Dictionary* and includes 213 verbs (Lönngren 1993). We used the data on these 213 verbs, and also performed on-line queries to select the experimental material for our lexical decision task experiment.

19. Strictly speaking, the term “L2 input” in this case is inaccurate, since the assessment of L2 input frequencies included both what the learners heard and produced themselves, in other words, both the input and the output.

20. The first figure corresponds to the number of verbs in the active vocabulary, and the second figure (in parentheses) to all the verbs from the active and passive vocabulary combined.

21. Both these corpora are available at the following url: http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html. The description of the Uppsala Corpus can be found in the Frequency Dictionary of Modern Russian (Lönngren 1993), which is actually based on this corpus. The description of the Tübingen corpus exists only in the electronic version, and can be accessed at: http://www.sfb441.uni-tuebingen.de/b1/en/korpora.html#interview.

At present, a new representative internet-based corpus of Russian is used to obtain both stem-cluster and whole-word frequencies (Sharoff 2002).[22]

4.2.2 *Material and method*

The series of experiments on Russian real and nonce verb generation by adult native and non-native speakers of Russian and Russian children compared the rates of stem recognition for the real and nonce verbs belonging to four different verb classes, the -aj-, -a-, -i-, and -ova- (see Tables 1 and 2 for the conjugational features and type frequencies of these classes).[23] Stem recognition, as a measure, does not count all the smaller errors in the application of the conjugational pattern, it is only concerned with the choice of one conjugational pattern over the other(s), e.g., -aj- versus -a-, etc. The -aj- class is the regular high type frequency default class representing the 'Vowel + J' conjugational pattern, which is characterized by no irregular morphophonemic processes, only the truncation of the 'J' before consonantal endings.[24] The -a- and -i- classes both represent the 'Vowel + Ø' pattern, which is characterized by some irregularities, such as consonant mutations and stress shifts.[25] At the same time, the -i- class is high type frequency productive, while the -a- class is low type frequency unproductive. The -ova- stem is regular, in a sense that all the forms in the paradigm are predictable, and productive, but it has a rare feature – suffix alternation – in its paradigm.

The token frequencies, or more exactly, stem-cluster frequencies of the verbs included in the testing material, which reflect the frequency of the stem in all the forms of a particular verb that occurred in the database, were obtained from *The Frequency Dictionary of Russian Language* by Zasorina (1977). This dictionary contains approximately 40,000 words and is based on a 1,000,000-word corpus of written Russian language including fiction, scientific texts, and newspaper and journal articles. The real verbs chosen for the experimental material belonged to two frequency ranges: very high-frequency and high-frequency, and the nonce verbs were created by manipulating the initial segments of the very high-frequency verbs (see Table 3). Note that the labels 'very high-frequency' and 'high-frequency' are relative, with the 'high-frequency' category corresponding rather to the mid-frequency range. The choice of the frequency ranges was dictated by the following consideration: For L1 children, both very high-frequency and high-frequency ranges should include the familiar verbs; for beginning L2 learners, only very high-frequency verbs should be familiar. The verbs were balanced in average token frequency across the classes. This produced 60 stimuli, to which 20 fillers were added, in two different versions of the experiment these were either an additional set of 20 nonce verbs, which paralleled

22. The corpus can be found at http://www.artint.ru/projects/frqlist/frqlist-en.asp.

23. In fact, these experiments are part of the on-going project, which includes other populations of speakers, such as children with language disabilities, patients with aphasia, and now schizophrenic patients. For the comparative data on normally developing children and children with language disabilities see Gor and Chernigovskaya 2003a.

24. In other words, its suffix includes a combination of one of the vowels and a 'J'.

25. In the 'Vowel + Ø' pattern, a single vowel constitutes the suffix.

Table 3. The testing material for the verb generation task

Verb Class/Infinitive/Basic Stem/Gloss	Very High-Frequency	High-Frequency	Nonce
-aj- gul'at' (gul'-aj-), "to walk"	N = 5 Average Fr. 173.2	N = 5 Average Fr. 10.8	N = 5
-a- pisat' (pis-a-), "to write"	N = 5 Average Fr. 174.4	N = 5 Average Fr. 13.4	N = 5
-i- nosit' (nos-i-), "to carry"	N = 5 Average Fr. 170.2	N = 5 Average Fr. 11.0	N = 5
-ova- probovat' (prob-ova-), "to try"	N = 5 Average Fr. 149.8	N = 5 Average Fr. 11.0	N = 5

the high-frequency verbs in the material, or 20 verbs belonging to other verb classes: -e-, -ej-, -avaj-, (i)j-, and (o)j-. The inclusion of a certain number of distractors from five more classes complicated the subjects' task by making the choice of the appropriate conjugational pattern less obvious and thus preventing the subjects from quickly developing response strategies. Two different versions of the experiment were created. In the first one the initial stimulus verb was presented to the subject in its infinitive form, while in the second one it was presented in the past-tense plural form. The subjects were required to generate the non-past 3rd person plural and 1st person singular forms of the verbal stimuli. All the verbs were embedded in simple carrying sentences, which together with follow-up questions formed a quasi-dialogue:

Infinitive

Experimenter: I want to __________.
Subject: Me too, I want to ________.
Experimenter: And what are you doing today?
Subject: Today I _______.[26]
Experimenter: And Mary and Peter?
Subject: Today they _______.

Past Tense

Experimenter: Yesterday they _______. And what are they doing today?
Subject: Today they _______.
Experimenter: And you?
Subject: Today I _______.

The task was performed orally, however, the adult subjects were provided the written script for each mini-dialog. Audio recordings of the sessions were further analyzed to establish the stem recognition and correct response rates. The verbs were presented in a quasi-random order with no two verbs belonging to the same stem following each other. Half of the subjects received the version of the text with all the stimuli presented in an inverse order to control for the possible effects of the order of presentation, such as fatigue or habituation to the task.

26. Russian uses a synthetic finite non-past tense verb form here.

The subjects for this set of experiments were:

Experiment 1A – 15 adult native speakers of Russian.[27]

Experiment 1B – 37 volunteer students at the University of Maryland at the end of their second semester of Russian in two groups of 20 and 17 subjects.[28]

Experiment 1C – 20 Russian children with normal language and cognitive development, and no hearing problems. There were 5 children aged 4, 9 children aged 5, and 6 children aged 6 in the group of subjects.

4.2.3 *Results and discussion*

This section will compare and discuss the data obtained in all the three experiments, 1A, 1B, and 1C. The results obtained for native Russian adults and American learners of Russian are represented in Figure 1.[29]

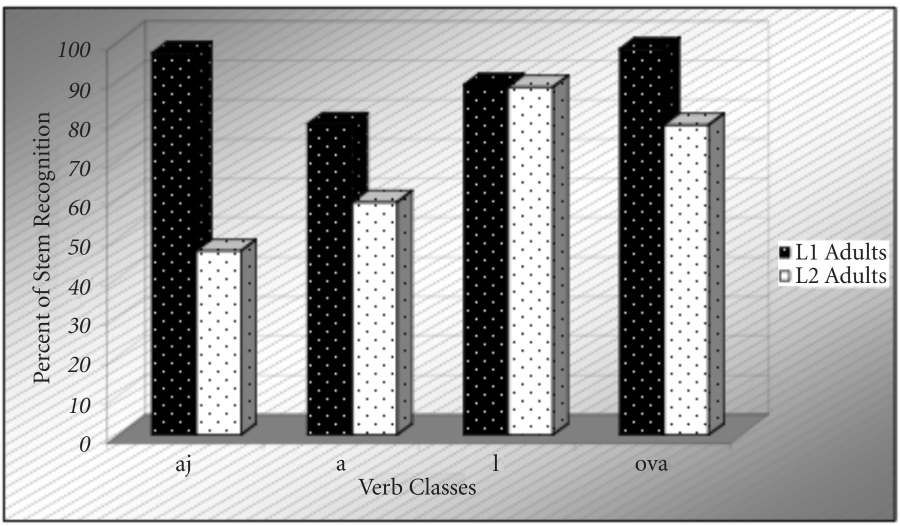


Figure 1. Rates of stem recognition in a verb generation task with adult L1 and L2 speakers (infinitive stimuli)

27. The data on adult L1 speakers were analyzed by Megan Malinowski and became part of her Master's Thesis.

28. Since the groups of 20 and 17 subjects were tested on different versions of the test, their data were not pooled together, but were compared to the matching data – 20 subjects were compared with 20 Russian children, while 17 subjects were compared with 15 Russian adults. The data on 17 L2 learners were analyzed by Natalia Romanova and became part of her Master's Thesis.

29. Since not all the adult and child L1 learners completed both versions of the test with the past-tense and infinitive stimuli, the comparisons will be made for the available data sets, either past-tense or infinitive.

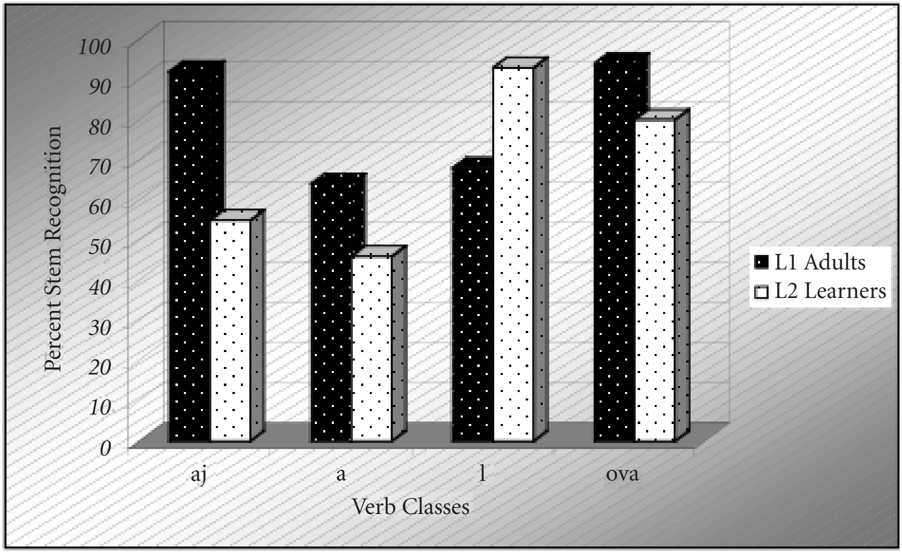


Figure 2. Rates of stem recognition for nonce verbs in L1 and L2 adult speakers (infinitive stimuli)

The rates of stem recognition for native Russian adults were predictably much higher than for L2 learners. Their data show the ceiling effect with the real infinitive stimuli, both very high- and high-frequency. In the averaged data, the rate of stem recognition was the lowest for the -a- class, since the infinitive is ambiguous due to stem consonant truncation, and the stem can be attributed either to the default -aj- class or the -a- class. The -i- class, which has a conflicting status – it is a high type frequency productive class, but has some irregularities in its conjugation – also caused some problems. The American learners, for whom most of the verbs, even some very high-frequency ones, were novel, showed much lower stem recognition rates on all the classes, except for the -i- class. What is remarkable, however, is that L2 learners were not able to rely on the default pattern when dealing with the 'ambiguous' infinitives of the -aj- and -a- classes. They chose one or the other pattern roughly equally often. This distribution reflects the composition of the experimental material – equal proportion of the -aj- and -a- verbs – rather than the type frequencies found in Russian.

Obviously, the comparison of the data on real and nonce verbs pooled together masks the results on novel verb generation, hence the necessity to isolate the nonce verb generation data (see Figure 2). Remarkably, while native speakers dealt with the novel -aj- and -a- verbs as expected based on type frequencies for these classes – they gave strong preference to the -aj- pattern – L2 learners did not show a strong preference, and overall, their stem recognition rates for these two 'ambiguous' stems were low. At the same time, they surpassed native speakers in processing the -i- verbs. This, again, shows that they were less influenced by the default 'Vowel + J' conjugational pattern. The -ova- class with suffix alternation was predictably more often recognized by L1 than L2 speakers given that the latter ones did not have sufficient exposure to that class.

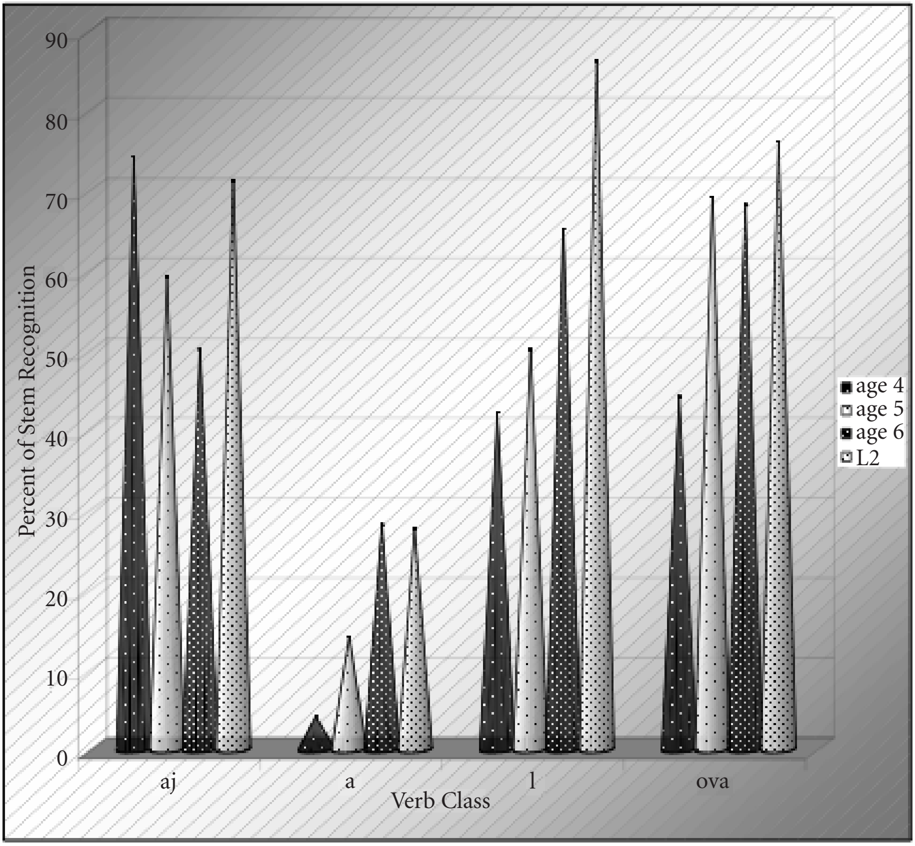


Figure 3. Rates of stem recognition for nonce verbs in L1 children and L2 adults (past-tense stimuli)

Figure 3 compares the data on verb generation in adult L2 learners and native Russian children broken down by the age. First, it clearly demonstrates a developmental tendency in L1 children, notably, they strongly rely on the default -aj- pattern at age 4, and then gradually move away from the default to the 'Vowel + Ø' pattern. Also, at age 5 the -ova- conjugational pattern stabilizes. The comparison of the child and adult L2 data demonstrates that L2 learners do not fit with any of the age groups: they perform like 4-year-olds on the -aj- verbs, like 6-year-olds on the -a- verbs, and outperform all the age groups on the -i- and -ova- verbs.

4.2.4 *Conclusion*

In conclusion, the verb generation experiment showed that the knowledge of probabilities influenced the processing of regular verbal morphology in all the three populations of speakers – L1 adults and children, as well as L2 adults. The generalization rates in novel verb generation depended on the type frequency, as well as the degree of regularity for different conjugational patterns. When dealing with ambiguous verbs, all the speakers preferred the high type frequency default conjugational pattern. At the same time, the processing of verbal morphology in L2 learners differed in substantial ways from that of adult native speakers – L2 learners were less biased towards the default than L1 speakers. One of the major differences between the type frequencies available to adult L1 and L2 speakers is the leveling of the size of verb classes in L2 compared to the full verbal system. For example, the -aj- class is roughly 12 times larger than the -a- class in Russian language, but only 3 times larger in the input to L2 learners (see Table 2). Therefore, the fact that L2 learners relied on default processing less than native speakers can be attributed in part to the reduced verbal system that is available to them. The study demonstrated that beginning L2 learners do not use the same probabilities as native speakers. Also, L2 rates of stem recognition did not match any of the child L1 age groups. At the same time, L1 children demonstrated a clear developmental tendency of moving away from the default pattern, which is acquired first. The Rules and Probabilities Model (Gor 2003, 2004) took into account these results, and also the results of another set of experiments with adult L1 and L2 speakers of Russian (Chernigovskaya & Gor 2000). The model claimed that both in L1 and L2 acquisition, and, more importantly, in endstate adult L1 processing of inflectional morphology, rules are applied with varying degrees of probabilities, the latter depending on input frequencies. The important part of the argumentation is the claim that the underlying mechanism in novel verb generation is based on rules and not phonological associations.[30] With regard to this set of native and non-native data, one can argue that all the speakers used the regular default 'Vowel + J' rule with different probabilities depending on the suffix vowel, the 'a' versus the 'i', in accordance with the type frequencies of the verb classes. In this sense, rules and probabilities interacted in adult and child L1 speakers and L2 learners. Certainly, only native adults could make use of the probabilities associated with the full Russian verbal system.

5. Lexical decision task

5.1 Lexical decision task with adult native speakers of English

Research on the processing of verbal morphology in English has been concerned with the question whether regularly inflected verbs are stored decomposed or in their inflected form. The most representative study, in fact, demonstrated that English regularly inflected verbs with the frequency above 6 per million are stored undecomposed, which deals a considerable blow to the claim of the dual-system approach that all regular verbs are inflected by symbolic rule computation, since the 6 per million threshold puts even less frequent inflected verb-forms on the list of stored items in English (Alegre & Gordon 1999). This comprehensive study had a broader focus and included a series of five lexical decision task (LDT) experiments investigating the effect of whole-word frequency of different classes of inflected words, not just past-tense verb forms, on reaction times (RTs). Thus, the inflected items included in the experiments consisted of plural nouns (*spheres, cellars*, etc.), and different inflected verb forms (*pretended, yelling, conveys*) with adjectives (*arrogant,*

30. See the details of this argumentation in Gor, 2004.

exquisite) representing uninflected words. As discussed above, this research paradigm is based on the assumption that sensitivity of RTs to whole-word or stem-cluster frequency can signal whether the inflected word is stored whole or decomposed into morphemes. Whole-word frequencies will influence RTs, if the inflected word is stored undecomposed. Stem-cluster frequencies will influence RTs if the word is stored decomposed, and the frequencies of all the inflected forms carrying the same stem are added together. Then, if the words are matched for stem-cluster frequency, but differ in whole-word frequency, and these differences have an effect on RTs in a LDT, one can argue that the inflected words from the experimental set are stored undecomposed. The experiments treated frequency as a continuous rather than a categorical variable and compared RTs to inflected and matched uninflected words. The experimental material in each of the 5 experiments was similar in composition: the target stimuli constituted one fourth of the testing material with fillers constituting three fourths of the material. For example, in Experiment 1, 66 experimental regularly inflected items were matched with 66 simple (uninflected) words, 66 "inflected" nonwords, and 66 simple nonwords (Alegre & Gordon 1999: 45–46).[31] The experimental procedure also remained the same across the five experiments. The LDT was programmed and performed using the Micro Experimental Lab (MEL) software on a PC environment. The main experiment was preceded by several practice trials; in the main part of the test, all the experimental and filler items were presented in a different random order to each subject. The participants received the instruction to press the "1" key on the keyboard if the stimulus appearing on the screen was a word, and "2" if it was a nonword. The stimuli were preceded by a fixation point, an asterisk, in the middle of the screen, and then appeared centered around the asterisk position for 1.5 seconds (s) or until the subject responded. All the slow answers which exceeded 1.5 s were discarded from further analyses. The experiments provided feedback to the participants, and when the answer was incorrect, the message "wrong answer" appeared on the screen.

The main part of experimental design was the establishment of stem-cluster and whole-word frequency ranges for each experiment depending on the hypothesis it tested, and finding the stimuli that belonged to the required range based on Francis and Kucera (1982). A caveat is in order when using this research method: Not only is it extremely difficult to find the verbs matched in stem-cluster frequency, but considerably differing in the whole-word frequency of a specific form in the paradigm, it is not always clear, which forms should be included in the count for stem-cluster frequency. For example, in Russian, a language with a developed conjugational paradigm, should one include all the four participles, two gerunds, and two imperative forms in addition to the non-past tense and past-tense forms in the verb's stem-cluster frequency count? The decision will considerably affect the frequency range.[32] Despite all these methodological difficulties,

---

31. The terms nonce word, pseudo-word, and nonword used by different researchers refer to the same type of word, an artificially created word with no meaning, which has some of the formal properties of existing words.

32. Note that even for English the issue of how to calculate verb stem-cluster frequencies is far from being transparent. For example, a study, which measured RTs in a verb generation task, did not include this research paradigm generates meaningful results. Thanks to a careful selection of the frequency ranges, and control for stem-cluster frequency of the stimuli varying in whole-word frequency, the Alegre and Gordon (1999) study produced the results necessary for the understanding of storage and retrieval of morphologically complex regular items. Some of their findings have direct bearing on the polemics surrounding English past-tense inflection. Thus, the study reliably demonstrated whole-word frequency effects above the threshold of 6 per million in regularly inflected words, which can be taken as an indication that they are stored undecomposed, as inflected word-forms, and not assembled on-line by the application of a symbolic rule, as predicted by the dual-system account.

5.2 Lexical decision task with adult native speakers of Russian

The computation of stem-cluster frequency remains a theoretical problem for Russian, a language with rich verbal morphology, both inflectional and derivational. Thus, it is not clear that all the participles and gerunds should be included in stem-cluster counts. However, in designing the LDT, we decided to pursue a different issue, and thus were able to avoid this methodological difficulty. The goal of the experiment was to prove that type frequency, or the frequency of the conjugational pattern, also affects RTs in a LDT. Combined with the argument that the conjugational pattern in question represents a symbolic rule rather than an analogy based on phonological similarity, this result would support the claim that rules have probabilities in adult native processing. In a series of previous studies we explored the role of type frequency as the frequency of the verb class (Chernigovskaya & Gor 2000; Gor & Chernigovskaya 2005). In the LDT, we wanted to test another hypothesis that different forms in the paradigm, which involve the use of a similar rule in terms of stem allomorphy, or in other words, display the same degree of regularity, but with different inflections, will produce different latencies in a LDT depending on the frequency of the form in the paradigm. Our analysis of the frequencies of the 213 most frequent Russian verbs confirmed the commonly known fact that the 3rd person singular verb form is the most frequent of the non-past tense forms, while the 2nd person singular is one of the least frequent forms. The infinitive has the frequency similar to the one found in the 3rd person singular, however, it is hypothesized to have a special status in terms of storage and retrieval, since it is the 'citation' form, and therefore, could have a faster access speed. Table 4 demonstrates that the distribution of frequencies discussed above holds for both corpora we were using.

The purpose of the LDT was to measure the speed of recognition of visually presented inflected verbs. The experiment tested the hypothesis that RTs depend on the type frequency of the forms in the conjugational paradigm. The subjects in the experiment see the words appearing on the screen and need to press two different buttons on the

---

the frequencies of regular past-tense forms in the stem-cluster frequency counts. This was done on the grounds that the frequencies of irregular past-tense verbs are not included in their stem-cluster frequency due to stem allomorphy, and the approach should be consistent for all the verbs. However, this led to a paradoxical situation when past-tense frequency could be higher than stem-cluster frequency (Prasada et al. 1990).

Table 4. Frequency distribution of the finite verb forms in the Uppsala and Tübingen Corpora (percentages based on the average for 213 most frequent verbs)

	Infinitive	**Non-Past Tense**					
Number		**Singular**			**Plural**		
Person		1st	2nd	3rd	1st	2nd	3rd
Uppsala Corpus	18.8	2.5	1.5	17.5	1.7	0.8	7.8
Tübingen Corpus	20.4	6.4	1.7	19.0	2.5	2.4	7.5

	Past Tense				**Imperative**	
	Masculine	**Feminine**	**Neuter**	**Plural**	**Singular**	**Plural**
Uppsala Corpus	20.9	10.1	6.0	10.7	1.1	0.5
Tübingen Corpus	14.6	7.2	5.3	10.8	0.6	1.0

button box or keys on the keyboard depending on whether they think that the word is a real word or non-word. For morphologically simple undecomposable words the speed of access would reflect their frequency, more frequent words will be accessed faster. For morphologically complex words the RT will reflect its whole-word frequency if it is accessed undecomposed, however, if it is decomposed in order to be accessed, the speed of decomposition would depend both on the frequency of the stem, which will be controlled by using the same verb in several inflected forms, and the frequency of the inflection rule or pattern. In order to make the LDT task reasonably challenging for the subjects, the experimenter needs to make the testing material sufficiently varied by including not only a matching number of non-words, but also numerous word and non-word fillers. Since the purpose of a LDT is to measure the latency of the first reaction to the visual verbal stimulus, the experiment needs to be fast-paced. This paper will discuss the methodology of the first version of the LDT experiment, LDT1, which was designed on a Mac platform using PsyScope software.[33] It will report the results both of the LDT1 experiment, and LDT2, a follow-up study designed for a PC platform using Presentation software.

5.2.1 *Testing material*

Both the pilot LDT1 and the follow-up LDT2 experiments used the same testing material: 5 high-frequency and 5 low-frequency verbs belonging to five different verb classes, the -aj-, -a-, -i-, -e-, and -ova- (see Tables 2 and 3 for their conjugational features and type frequencies). The -aj- class is the largest regular, productive, and default class; the -a- class is a much smaller unproductive class with a predictable but rare pattern of stem allomorphy; the infinitives of these two classes are indistinguishable and both end in *-at'*. The

33. A more detailed account of this experiment can be found in Gor 2003. Many thanks to Natalia Sliusar for help with the preparation of the testing material and design of the experiment with PsyScope.

-i- and -e- classes have the same conjugational patterns involving the stem allomorphy similar to the one found in the -a- class, but distributed differently across the paradigm; however, the -i- class is large and productive, while the -e- class is small and unproductive. And finally, the -ova- class is large and productive, and has suffix alternation in its paradigm. According to the dual-system approach criteria, the -aj-, -i-, and -ova- classes are regular, because they are productive. The high-frequency verbs belonged to the 10–25 per million frequency range as found in the Frequency Dictionary (Lönngren 1993) based on the Uppsala Corpus (this count reflects the stem-cluster frequency), while the low-frequency verbs were not found in the dictionary at all, and therefore had the stem-cluster frequency of less than 1 per million.[34] We used three forms of each verb, the infinitive, 2nd person singular, and 3rd person singular, and a set of matching nonce verbs, created by manipulating the initial segments of the real verbs used in the experiment, usually, the initial consonant. Thus, we included 150 real (5 verbs × 3 forms × 5 classes × 2 frequency ranges) and 150 matching nonce verbs. Since it is impossible to match the verbs belonging to the five classes exactly in length given that the length of their suffixes differs, the experiment controlled for the following factors: all the verbs were prefixless and belonged to the imperfective aspect. Their length varied from 2 to 3 syllables and from 7 to 9 letters. The fillers were 15 high-frequency and 15 low-frequency nouns matched in stem-cluster frequency and length with the verbs used in three cases, the Nominative singular (the citation form), Dative plural, and Instrumental singular. For each of the real nouns a matching nonce noun was created. The number of nouns was 90 (15 nouns × 3 forms × 2 frequency ranges) plus 90 matching nonce nouns, thus bringing the total number of stimuli in the experiment to 480.

5.2.2 *Method*

The presentation of the stimuli to the subjects and the measurement and recording of the RTs were controlled using PsyScope, a software designed for the Mac environment. PsyScope eliminates the programming stage thus allowing the researcher to work within the graphic environment, and visually represents the experimental design. It was developed by a team of researchers at Carnegie Mellon University and made available for research purposes through the site at http://psyscope.psy.cmu.edu/.[35] One of the factors we needed to control for was the number of repetitions of the stem. Remember that the subject was exposed to three different word-forms of the same verb or noun, and the RTs can be expected to decrease with each repetition of the same lexeme within the same experiment. This is a well-attested effect observed in repetition priming studies. We dealt with this problem in the following way: The total number of 480 stimuli was divided in three

34. As already noted above, the labels for frequency ranges do not carry any absolute values, in fact, the high-frequency category includes the mid-frequency range, while the low-frequency category includes very low-frequency verbs.

35. A detailed manual can be downloaded from this site, which takes the beginning researcher through all the necessary steps to set up the experiment. At present, PsyScope does not work with Mac OS X, but it is currently being updated to be fully compatible with the new operating system.

blocks with 160 items in each, and the material was balanced across the blocks with the form of each word occurring only once within the block. Then, we prepared six different versions of the test with all the possible sequences of the blocks, and presented a different version of the test to each of the participants in groups of six. With this design, the repetition effects remained present for each individual subject, but were counterbalanced across the subjects. A practice session with 12 stimuli preceded the main three blocks. The stimuli appeared on the screen in random order and disappeared after 600 milliseconds (ms). The subject pressed one of the buttons on the button box for a word and another for a nonword, at which moment the inter-stimulus interval was initiated.[36] This interval varied in duration between 700 and 1100 ms to avoid the effect of habituation to the fixed interval when the subject shows a better readiness towards the end of the trial. If the subject failed to press any button, after 6000 ms the inter-stimulus interval was initiated automatically. PsyScope records the RTs and the type of response (the word/nonword decision) in a chart that is easily imported into another program, such as Excel or SPSS to be further analyzed. The first pilot series was conducted with 8 volunteers, adult native speakers of Russian.

5.2.3 *Results and discussion*

The RTs of each of the 8 subjects were averaged to see if there were any outliers in the group, and indeed, the RTs of one participant were considerably longer than those of the other participants (see Figure 4), and the data for this participant were eliminated from further analyses.

The first result of this pilot study was the confirmation of the expected effect of word frequency on response latencies, since indeed, the high-frequency verbs produced shorter RTs than the low-frequency verbs across all the five verb classes (see Figure 5).

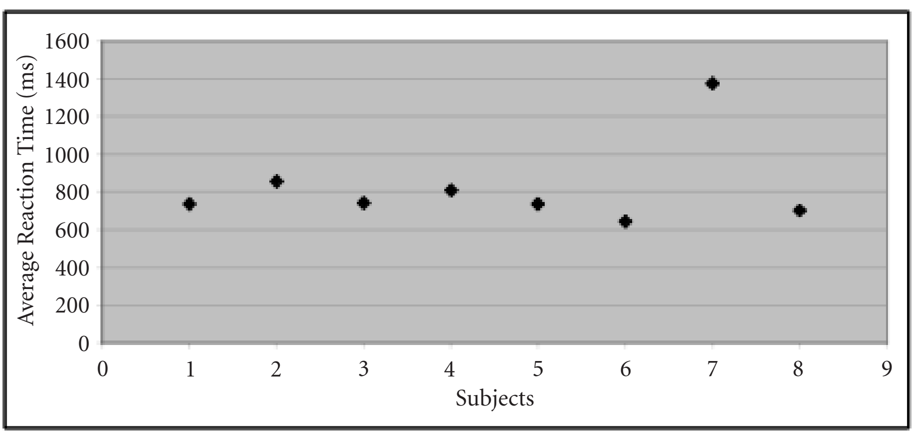


Figure 4. Average RTs of individual participants on a lexical decision task

36. The use of the button box makes it possible to measure RTs with the accuracy of up to 1 ms.

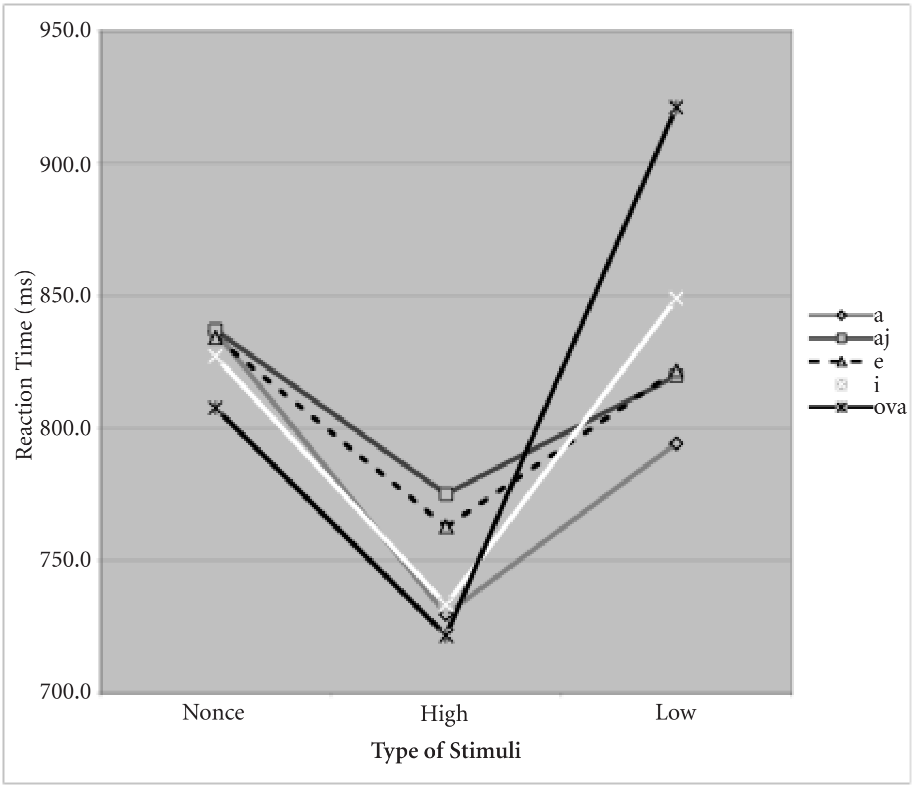


Figure 5. Reaction times to high-frequency, low-frequency, and nonce verbs

The role of type frequency was most evident in high-frequency verbs, where the difference between the latencies found in the 2nd person singular, 3rd person singular, and the infinitive is the greatest (see Figure 6).

Given the frequency range used in the experiment, one can expect type frequency to affect the RTs to a lesser degree in low-frequency verbs where all the forms in the paradigm have the frequency of less than 1 per million. In nonce verbs, the infinitive required slightly more time to process than the 3rd person singular form, but the difference between the 2nd person singular and the 3rd person singular was maintained. At the same time, the LDT1 experiment showed no effect of the verb class on RTs in visual perception of inflected verbs.

This experiment was replicated at Saint Petersburg State University, Russia using Presentation software on a PC platform (Sliusar 2003).[37] In the modified version, LDT2, the testing material was the same as in LDT1. A group of 28 adult native speakers of Russian took part in this experiment, 16 women and 12 men aged 15–50. The analysis of variance demonstrated that for all the stimuli combined (high-, low-frequency, and nonce

37. The modified LDT2 experiment was part of the Master's Thesis by Natalia Sliusar.

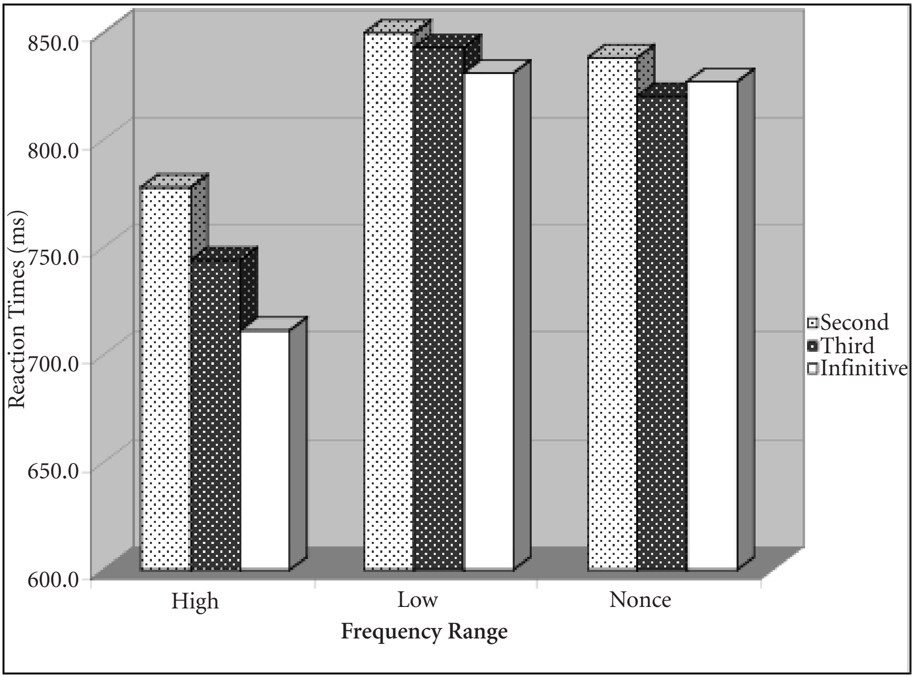


Figure 6. Reaction times to three different forms in the paradigm

verbs), four out of the five tested factors were statistically significant – frequency, number of repetitions, form in the paradigm, and length in letters (Sliusar 2003: 71 ff.). The verb class factor was again not significant. The analysis of the three frequency ranges separately revealed that the number of repetitions was a highly important factor in the nonce verb category, and to a lesser degree, in the low-frequency category. The length in letters factor emerged only in the low-frequency category, while the form in the paradigm (2nd versus 3rd person singular versus the infinitive) factor became highly significant only in the high-frequency category.

To summarize the results of both versions of the LDT, the type frequency of the form in the paradigm is a significant factor affecting RTs for the verbs with stem-cluster frequency of 10 per mln and above (the high-frequency category in the experiment). Since 2nd person forms have lower type frequency in speech and take longer to process, one can conclude that type frequency affects the recognition of inflected words. The role of the other factors, such as word frequency, the number of repetitions, and length in letters, is not a new finding in a LDT, but their significance in the present study validates the other finding. Let us now address briefly the kind of processes involved in verb recognition in our LDT. The issue in question is whether inflected forms are stored in the mental lexicon undecomposed as whole words, or decomposed into stems and inflections. The lexical decision about the verb stimuli will be based on whole word access in the first case and decomposition in the second. One can argue based on the average frequencies for the forms in the paradigm calculated for the 213 most frequent verbs in the Uppsala and Tübingen

Corpora (see Table 4), that the infinitives and 3rd person singular forms each account for approximately 20% of the stem-cluster frequency. Therefore, for the verbs with stem-cluster frequency of 10–25 per mln, the frequency of the infinitive and 3rd person singular will be roughly in the 2–5 per mln range. The 2nd person singular will be 10 times less frequent, and therefore will fall below 1 per mln, as well as all the frequencies of the forms of low-frequency verbs with stem-cluster frequency already less than 1 per mln. Based on the comparison of the estimated whole-word frequencies of the Russian forms in the paradigm chosen for the LDT with the 6 per million threshold obtained by Alegre and Gordon (1999), it appears that all the word-forms included in the experiment are below the threshold for English. It is important to remember that the threshold found for English, with 7 per million being the lowest frequency showing the whole-word effects, may not be relevant for Russian. And indeed, Russian has so many more forms in the conjugational paradigm than English that a threshold on storage similar to English would imply that there are several times more verb forms stored in the Russian mental lexicon. While in principle this is not impossible, it still needs empirical proof, which is lacking at present. In the absence of the data indicating that Russian has a larger storage space for the inflected forms, it is safe to assume that all the verb forms in our LDT are stored decomposed. This would mean that in order to perform the LDT, the subjects would need to analyze and decompose the stimuli. This operation took less time with the higher-frequency conjugational pattern, with the differences most visible in the higher token frequency range. However, the infinitive produced shorter response latencies, than the 3rd person singular, which indicates that it may be stored and accessed differently, possibly, as a whole word, given its special status of a citation form.

5.2.4 *Conclusion*

The results of the LDT experiment confirmed the hypothesis that the frequency of the form in the paradigm affects the response latency to a visually presented verb. While the experiment was not designed to determine whether the verb forms are stored decomposed or as a whole word and which frequency ranges correspond to these two different kinds of storage, it clearly demonstrated that 2nd person singular non-past tense forms generate longer RTs than 3rd person singular forms, and the RTs to the infinitive are the shortest. The fact that 2nd person singular forms are much more rare and at the same time take more time to process, and this effect remains across the verb classes, supports the idea that the frequency of the pattern or rule plays a role in morphological processing. Combined with a claim that Russian verb conjugation is rule-based rather than phonological association-based advocated elsewhere (see Gor 2004), these results support the hypothesis that rule application is affected by the probability of the verb form, which involves the use of this rule. The obtained results raise the issue of the unit of storage – whether it is the basic stem, or the infinitive, or both, or something else, and whether this depends on the regularity of the morphological processes. The results of the pilot LDT1 experiment indicate that the infinitive is recognized faster than the 3rd person singular form, despite the fact that these two forms have very similar estimated whole-word frequencies. This difference is smaller in the low-frequency verbs, but it does not disappear. At the same time, this tendency is absent in nonce verbs. These facts support the idea that the infinitive has a special status with regards to storage, hence no shorter latency in nonce verbs, in which none of the forms can be potentially stored.[38]

6. General conclusions

Two experimental methods, a real and nonce verb generation task and a lexical decision task reviewed in this paper, made significant contributions to the understanding of the mechanisms underlying the processing of inflectional morphology, and more broadly, linguistic processing in general. The use of these experimental methods is enriched by targeting different populations of speakers, native adults and children, as well as second language learners, and also by focusing on languages with complex inflectional morphology and numerous verb classes varying in the degree of regularity, in addition to English. The reviewed studies with Italian and Russian verbs demonstrated that phonological similarity and type frequency influenced generalization rates in verb generation experiments. The results of LDT experiments with Russian stimuli showed that both type and token frequencies affect RTs. Also, a LDT experiment with English word stimuli provided evidence that the whole-word representations of morphologically complex inflected words, including inflected verbs, with the frequency above 6 per million are available in linguistic processing. The latter finding challenges the dual-system account of morphological processing. The observed influence of the input frequencies suggests that the speakers brought to those two tasks, the verb generation task and the LDT, their knowledge of linguistic probabilities. The Russian data demonstrated that, while token frequency influenced the speed of access to or retrieval of the lexical item, type frequency, or the frequency of the pattern, influenced generalization rates in novel verb generation and RTs in a LDT with visually presented verbal stimuli. In the verb generation task, it was the frequency of the verb class that affected the generalization rates. In the LDT, the frequency of the particular form in the paradigm was one of the significant factors for the high-frequency verbs. The presence of these type frequency effects across all the verb classes, including the regular default class, implies that the kind of processing, which would qualify as regular according to the dual-system approach standards, is influenced by linguistic probabilities. In this sense, the processing of Russian inflectional morphology exhibits the influence of probabilities on rule application. At the same time, the two sets of studies showed the differences in verbal processing in three populations of speakers – native Russian adults and children, and American learners of Russian. Only adult L1 speakers could fully access the frequencies observed in native Russian and make efficient use of the native probabilities. Adult beginning L2 learners' performance on the verb generation task reflected the fact that the frequencies they relied on were non-native, as they were derived from a reduced verbal system available to them. However, adult L2 stem recognition rates in nonce verb generation were in most cases higher than those of L1 children. This last population of

38. Unfortunately, no post hoc paired comparisons tests were reported in the Sliusar paper.

speakers, L1 children, demonstrated a clear developmental tendency from age 4 to 6 in a verb generation task.

And finally, experimental study of the processing of complex verbal morphology in several languages challenges the original strong claims made by the proponents of two main theoretical positions, the dual-system and the single-system approaches. More precisely, it shows that the "litmus test" used by both parties, which relies on the assumption that input frequencies and probabilities, as well as phonological similarity play a role only in associative patterning, and not symbolic rule computation, turns out to be false. In fact, the results of the reported studies suggest that frequencies and phonological similarity influence both rule-based and association-based processing, and therefore the frequency argument cannot be used to qualify the type of processing as associative patterning or symbolic rule application. At the same time, in addition to this negative result, the research data obtained by different experimental methods suggest another possibility, namely, that symbolic rule computation is not completely immune to the influence of linguistic probabilities, and that there is no categorical division between the lexicon and grammar, on the opposite, these two entities are interrelated. In a sense, this development would illustrate the position expressed in the Minimalist Program that words are inflected in the lexicon (Chomsky 1995).

References

Alegre, M. & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language, 40*, 41–61.

Bod, R., Hay, J., & Jannedy, S. (2003). *Probabilistic linguistics*. Cambridge: MIT Press.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form.* Philadelphia: John Benjamins Publishers.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*, 425–455.

Bybee, J. L. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition, 24*, 215–221.

Bybee, J. L. & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language, 58*, 265–289.

Chernigovskaya T. & Gor, K. (2000). The complexity of paradigm and input frequencies in native and second language verbal processing: Evidence from Russian. *Language and Speech Behavior*, *3(II)*, 20–37.

Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.

Clahsen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences, 22*, 991–1060.

Davidson, D. E., Gor, K. S., & Lekic, M. D. (1996). *Russian: Stage one: Live from Moscow!* Dubuque, Iowa: Kendall/Hunt Publishing Company.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, M. D. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge: MIT Press.

Francis, W. N. & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston, MA: Houghton, Mifflin.

Gor, K. (2003). Symbolic rules versus analogy in the processing of complex verbal morphology. *Regards Croisés sur L'Analogie. Revue d'Intelligence Artificielle*, *17(5–6)*, 823–840.

Gor, K. (2004). The rules and probabilities model of native and second language morphological processing. In L. Verbitskaya, T. Chernigovskaya (Eds.), *Theoretical problems in linguistics. Papers dedicated to 140 Anniversary of the Department of General Linguistics, St. Petersburg State University* (pp. 51–75). St. Petersburg: Philological Faculty of St. Petersburg State University Press.

Gor, K. & Chernigovskaya, T. (2001). Rules in the processing of Russian verbal morphology. In G. Zybatow, U. Junghanns, G. Mehlhorn, & L. Szucsich (Eds.), *Current issues in formal Slavic linguistics* (pp. 528–535). Frankfurt/Main: Peter Lang.

Gor, K. & Chernigovskaya, T. (2003a). Mental lexicon structure in L1 and L2 acquisition: Evidence from Russian. *Glossos*, *4*. http://www.seelrc.org/glossos/issues/4/.

Gor, K. & Chernigovskaya, T. (2003b). Generation of complex verbal morphology in first and second language acquisition: Evidence from Russian. *Nordlyd, 31*(6), 819–833. http://www.ub.uit.no/munin/nordlyd/.

Gor, K. & Chernigovskaya, T. (2005). Formal instruction and the mental lexicon: The acquisition of verbal morphology. In A. Housen & P. Pierrard (Eds.), *Investigations in instructed second language acquisition* (pp. 103–136). Berlin, New York: Mouton De Gruyter.

Jaeger, J. J., Lockwood, A. H., Kemmerer, D. L., Van Valin, Jr. R. D., Murphy, B. W., & Khalak, H. G. (1996). A positron emission tomographic study of regular and irregular verb morphology in English. *Language, 72*, 451–497.

Jakobson, R. O. (1948). Russian conjugation. *Word, 4*, 155–167.

Langacker, R. (1987). *Foundations of cognitive grammar, Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. (1988). A usage-based model. In B. Rudzka-Ostyn (Eds.), *Topics in cognitive linguistics* (pp. 127–161). Amsterdam: John Benjamins.

Lönngren, L. (Ed.). (1993). *A frequency dictionary of modern Russian: With a summary in English*. Uppsala.

MacWhinney, B. & Leinbach, J. (1991). Implementations are not conceptualizations: Revisiting the verb learning model. *Cognition, 40*, 121–157.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (1992). *Overregularization in language acquisition. Monographs of the Society for Research in Child Development, 57*(4), Serial No. 228. Chicago: University of Chicago Press.

Marcus, G. F., Brinkmann, U., Clahsen, H., Weise, R. & Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology, 29*, 189–256.

Matcovich, O. (1998). Regular inflection in the mental lexicon: Evidence from Italian. *Proceedings from the conference "The Verb in Cognitive Grammar"*. Gran.

Orsolini, M. & Marslen-Wilson, W. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes, 12*, 1–47.

Orsolini, M., Fanari, R., & Bowles, H. (1998). Acquiring regular and irregular inflection in a language with verb classes. *Language and Cognitive Processes, 13*, 425–464.

Pinker, S. (1991). Rules of language. *Science, 253*, 530–535.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York, NY: Basic Books.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*, 73–193.

Pinker, S. & Prince, A. (1991). Regular and irregular morphology in the psychological status of grammar. *Proceedings of the 1991 meeting of the Berkeley Linguistics Society.* Berkeley, CA: University of California.

Pinker, S. & Prince, A. (1994). Regular and irregular morphology and the psychological status of rules of grammar. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.), *The reality of linguistic rules* (pp. 353–388). Amsterdam: John Benjamins Publishers.

Pinker, S. & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences, 6*, 456–463.

Plunkett, K. & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition, 38*, 43–102.

Plunkett, K. & Marchman V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.

Prasada, S., Pinker, S., & Snyder, W. (1990). Some evidence that irregular forms are retrieved from memory but regular forms are rule-generated. Paper presented at *The 31st Annual Meeting of the Psychonomics Society*, 16–18 November.

Prasada, S. & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes, 8*, 1–56.

Ragnasdóttir H., Simonsen, H. G., & Plunkett, K. (1997). Acquisition of past tense inflection in Icelandic and Norwegian Children. In E. V. Clark (Ed.), *Proceedings of the 28th Annual Child Language Research Forum*. Stanford.

Rumelhart, D. E. & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructures of cognition*, Vol. 2 (pp. 216–271). Cambridge, MA: Bradford/MIT Press.

Sharoff, S. (2002). *The frequency dictionary for Russian*. http://www.artint.ru/projects/frqlist/frqlist-en.asp

Simonsen, H. G. (2000). Past tense acquisition and processing in Norwegian: Experimental evidence. *Language and Language Behavior 3(II)*, 86–101.

Sliusar, N. (2003). *A psycholinguistic study of the structure of the mental lexicon based on the Russian verbs*. M.A. Thesis, St. Petersburg: St. Petersburg State University.

Townsend, C. E. (1975). *Russian Word-Formation*. Slavica Publishers.

Ullman, M. T. (1999). Acceptability ratings of regular and irregular past-tense forms: Evidence for a dual-system model of language from word frequency and phonological neighborhood effects. *Language and Cognitive Processes, 14*, 47–67.

Ullman, M. (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilihgualism, 4*, 105–122.

Yang, C. D. (2002). *Knowledge and learning in natural language.* Oxford University Press.

Zalizniak, A. A. (1980). *The grammatical dictionary of Russian language*. Moscow: Russian Language Publishers.

Zasorina, L. N. (Ed.). (1977). *The frequency dictionary of Russian language*. Moscow: Russian Language Publishers.

Electrifying results

ERP data and cognitive linguistics

Seana Coulson

1. Introduction

Have you ever heard anyone say that people typically use only 10% of their brains? I have. Interestingly, I've never heard anyone suggest that we typically only use 10% of our stomach, 10% of our kidneys, or 10% of our heart. Yet, like the stomach, the kidneys, and the heart, the brain is an organ made up of cells that undergo continuous metabolic activity. One thing that surely contributes to the "10%" myth is the way that neuroimaging data are shown pictorially. In arguing for the notion that people only use a small portion of their brains at a time, a friend of mine cited "studies" by "scientists" that show that only small areas of the brain light up. In such studies, PET (positron emission tomography) or fMRI (functional magnetic resonance imaging) is used to monitor changes in metabolic activity, such as blood flow or oxygen consumption, in different brain regions as a function of doing various cognitive tasks. Inspection of the raw images suggests that most of the brain is, in fact, active at any given moment. Neuroscientists, however, are most interested in the differences in brain activity as a function of task (e.g. reading words as opposed to tapping one's fingers up and down). Consequently, the pictures show only those brain areas that are *more* active during the performance of one task than the other.

A more direct way of measuring brain activity is to record electrical activity in its cells, also called *neurons*. The primary mechanism of neural computation is the action potential, or spike. A spike is a sudden change in the cell's voltage due to an influx of sodium ions followed by an expulsion of potassium. When a neuron is stimulated, its firing rate (how often spikes occur) changes, usually by getting faster. However, neurons fire even when they are not being stimulated, something known as their baseline firing rate. The cells in the brain, then, are always active in the sense that spikes occur at a particular frequency, and the firing rate changes as a function of the cell's engagement in some brain activity. Moreover, besides spikes, neurons receive a constant barrage of chemical signals from one another that result in slow changes to their voltage known as post-synaptic potentials.

As noted above, the brain is an organ which, like the stomach, the kidneys, and the heart, is made up of cells. Unlike the stomach, the kidneys, and the heart, however, the activity of cells in the brain is intimately related to our ability to see, hear, feel, think, and use language. One way to study the relationship between electrical activity in the brain and language comprehension is to record event-related brain potentials (ERPs). Rather than placing electrodes directly into the brain, ERPs are recorded by applying electrodes to the scalp, usually by having volunteers wear a hat that looks something like a swimming cap with electrodes embedded in it. This non-invasive method can nonetheless provide us with a direct index of brain activity. ERPs represent electrical activity in the brain that is time-locked to the onset of a cognitive or motor event, and are known to be sensitive to many of the processing operations involved in the production and comprehension of language.

This chapter reviews methods and data in the domain of electrophysiology of language comprehension. We begin with a general introduction to the electroencephalogram (EEG) and event-related brain potentials (ERPs), and give an overview of language-sensitive ERP components. Section 4 reviews the way that ERPs have previously been used to address issues in cognitive linguistics, and Section 5 points to a number of ways that this technique could be further employed by cognitive linguists.

2. EEG and ERPs

Work on the cognitive neuroscience of language has attempted to monitor how the brain changes with manipulations of particular linguistic representations. The assumption is that different language sub-processes are subserved by different anatomical and physiological substrates that will generate distinct patterns of biological activity. These patterns can then be detected by methods sensitive to electromagnetic activity in the brain, such as the electroencephalograph, or EEG. An EEG is a non-invasive measure of physiological activity in the brain made by hooking up electrodes to the subject's scalp. These electrodes pick up electrical signals naturally produced by the brain and transmit them to bioamplifiers. Early versions of the EEG used a galvanometer (an instrument that detects the direction of small electrical currents) to move a pen on a rolling piece of paper. In more modern EEG systems, the bioamplifiers convert information about voltage changes on the scalp to a digital signal that can be stored on a computer.

The idea of using EEG to study the brain processes that support the complex activity of language comprehension is frequently met with a healthy dose of skepticism. After all, it does seem far-fetched to assume that electrodes sitting on the surface of the head could be used to discern activity in the participants' brains. Initially, it might seem a bit like trying to study how a car works by listening to it with a stethoscope. However, although a functioning engine produces sound, the sound is not an integral part of its function. In contrast, the electrical activity in the brain *is* an integral part of its function. Neurons in different parts of the brain communicate with one another via electrochemical signals, and the success or failure of any given cognitive activity can depend on whether a particular pattern of firing occurs in the brain.

2.1 EEG

Because the brain constantly generates electrical activity, electrodes placed on the scalp can be used to record the electrical activity of the cortex (the outer part of the brain, and the part thought to be most integrally involved in cognition). The EEG amplifies tiny electrical potentials and records them in patterns called brain waves. Brain waves vary according to a person's state, as patterns observed when a person is alert and mentally active differ from those observed when she is relaxed and calm, and those observed when she is sleeping. Like any wave – an ocean wave, a sound wave, or a light wave – a brainwave can be described by certain mathematical properties. For example, it has a crest (or positive peak) and a trough (or negative peak). At any given point in a wave, one can ask what its *amplitude*, or size, is (see Figure 1). Typically, when one refers to the amplitude of a wave, it's the size of the wave at the peak that's being described. The *wavelength* of a wave is the distance between a point on a wave and the corresponding point on the next wave in the wave train, such as between the crest of one wave and the crest of the next. Finally, the *frequency* of a wave describes how many waves (or cycles) are made per second. When describing brainwaves, most neuroscientists focus on the amplitude and the frequency, especially since frequency and wavelength are two different ways of describing the same thing (see Figure 1).

The pattern of electrical activity in the fully awake person is a mixture of brainwaves that occur at many different frequencies. However, the EEG is dominated by waves of relatively fast frequencies between 15 and 20 cycles per second (or Hertz), referred to as beta activity. If the subject relaxes and closes her eyes, a distinctive pattern known as the alpha rhythm appears. The alpha rhythm consists of brainwaves oscillating at a frequency

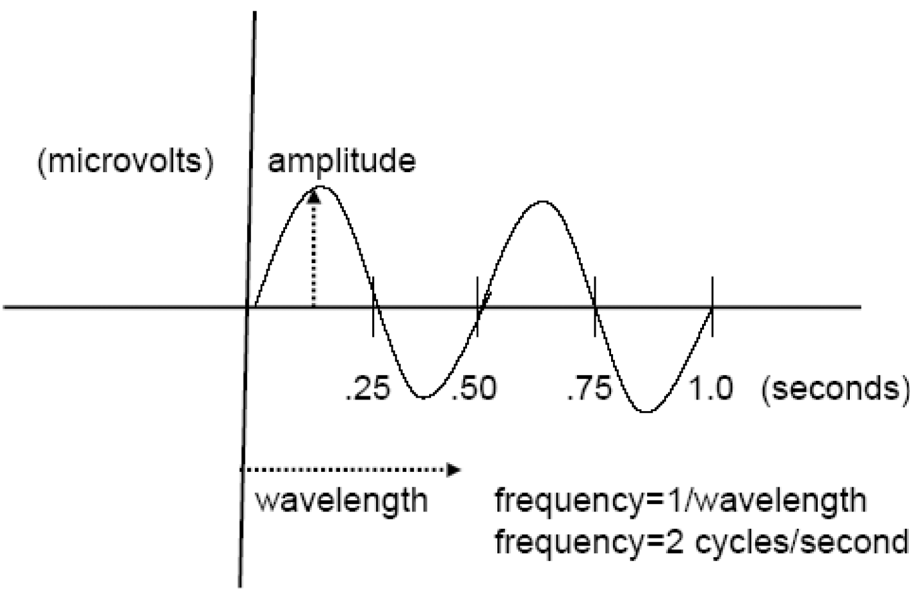


Figure 1. Two cycles of a waveform extended in time

that ranges from 9–12 Hertz. As the subject falls asleep, her brain waves will begin to include large-amplitude delta waves at a frequency of 1 Hertz.

Although the EEG provides overall information about a person's mental state, it can tell us little about the brain's response to specific stimuli. This is because there is so much background activity in the form of spontaneous brain waves it is difficult to identify which brain wave changes are related to the brain's processing of a specific stimulus and which are related to the many on-going neural processes occurring at any given time. In order to better isolate the information in the EEG that is associated with specific processing events, cognitive electrophysiologists average EEG that is time-locked to the onset of particular sorts of stimuli, or to the initiation of a motor response. The average EEG signal obtained in this way is known as the event-related potential, or ERP.

2.2 ERPs

ERPs are patterned voltage changes in the on-going EEG that are time-locked to classes of specific processing events. Most commonly these events involve the onset of stimuli, but they can also include the execution of a motor response (Hillyard & Kutas 1983; Rugg & Coles 1995). As noted above, we obtain ERPs by recording subjects' EEG and averaging the brain response to stimulus events. The effect of averaging different numbers of EEG trials can be seen in Figure 2. Each graph in Figure 2 shows the averaged EEG signal from an electrode at the top of the head that was recorded as a healthy adult listened to piano chords. Time is on the x-axis and voltage is on the y-axis, and thus the graphs show how voltage changes as a function of time. By convention, negative voltage is plotted upwards (not downwards).

The logic behind averaging, of course, is to extract from the EEG only that information which is time-locked to the processing of the event. The graph at the top of Figure 2 depicts the average signal from 3 trials. That is, for each time point measured (which in this case is every 4 milliseconds), the average voltage from the 3 trials has been plotted. This waveform has the oscillatory characteristics of the EEG and shows a substantial amount of activity during the time before the presentation of the chord (time 0 marks the onset of the stimulus). This is because the signal reflects not only the processing of the chord but brain activity related to other processes as well.

The different graphs in Figure 2 show how the signal changes when additional EEG trials are averaged into the waveform. The assumption is that brain activity related to the processing of the chord will be timelocked to its presentation, while brain activity related to other processes will not. The more trials in the averaged waveform, the more likely it is that the background activity will average to zero, so that the signal exclusively reflects the processing of the stimulus. Thus the graph at the bottom of Figure 2 shows virtually no activity before the presentation of the stimulus, followed by a series of peaks (the N1-P2 complex) that resemble those seen in other experiments in which people are presented with auditory stimuli.

Cognitive neuroscientists refer to the averaged signal as the event-related potential (ERP) because it represents electrical activity in the brain associated with the processing of a given class of events. In Figure 2, the "event" was the processing of a piano chord.

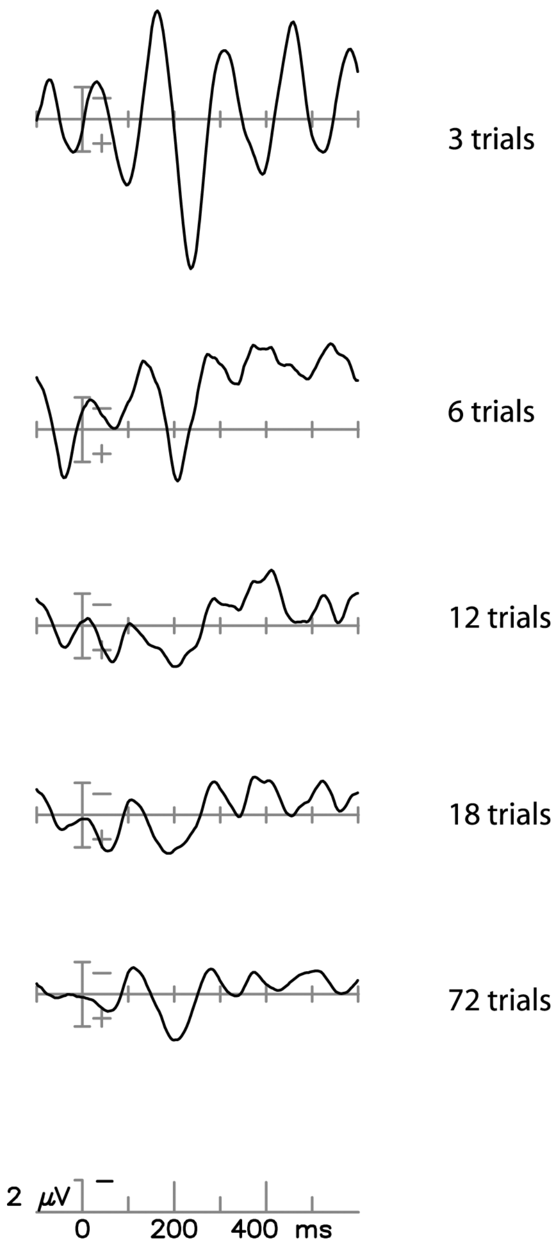


Figure 2. ERPs to piano chords from different numbers of trials (3, 6, 12, 18, and 72). Increasing the number of trials in the average decreases the amount of background activity and better represents activity related to processing the stimulus.

ERP language research involves interesting linguistic events such as the presentation of particular sorts of words. For example, in early work on language processing, Kutas and Hillyard (Kutas & Hillyard 1980) recorded ERPs to the last word of sentences that either ended congruously (as in (1)), or incongruously (as in (2)).

(1) I take my coffee with cream and sugar.

(2) I take my coffee with cream and dog.

Because the EEG associated with the presentation of a single event is relatively inscrutable, Kutas & Hillyard constructed 70 sentences, half of which ended congruously, half incongruously (Kutas & Hillyard 1980). By averaging the signal elicited by congruous and incongruous sentence completions, respectively, these investigators were able to reveal systematic differences in the brain's electrical response to these stimulus categories in a particular portion of the ERP that they referred to as the N400 component. Subsequent research has shown that N400 components are generated whenever stimulus events involve meaningful processing of the stimuli, and that its size is sensitive to fairly subtle differences in the processing difficulty of the words that elicit it. As such, many investigators have used the N400 component of the brain waves as a dependent variable in psycholinguistic experiments (see Kutas, Federmeier, Coulson, King, & Muente 2000 for review).

2.3 ERP components

While EEG measures spontaneous activity of the brain and is primarily characterized by rhythmic electrical activity, the ERP is a waveform containing a series of deflections that appear to the eye as positive and negative peaks. Such peaks are often referred to as components, and much of cognitive electrophysiology has been directed at establishing their functional significance. ERP components are characterized by their *polarity*, that is, whether they are positive- or negative- going, their *latency*, the time point where the component reaches its largest amplitude, and their *scalp distribution*, or the pattern of relative amplitudes the component has across all recording sites. The N400, for instance, is a negative-going wave that peaks approximately 400 ms after the onset of the stimulus, and has a centro-parietal distribution (evident over the back of the head) which is slightly larger over the right hemisphere.

The ERP approach seeks correlations between the dimensions of ERP components elicited by different stimuli and putatively relevant dimensions of the stimuli themselves. ERP components with latencies under 100 ms are highly sensitive to systematic variations in the physical parameters of the evoking stimulus. Because their amplitudes and latencies seem to be determined by factors outside the subject, they are referred to as *exogenous* components. In contrast, *endogenous* ERP components are less sensitive to physical aspects of the stimulus, reflecting instead the psychological state of the subject. Endogenous components are modulated by task demands and other manipulations that affect the subjects' expectancies, strategies, and mental set. Because they are thought to reflect cognitive rather than perceptual processing, most language researchers have restricted their attention to the endogenous ERP components elicited by linguistic stimuli.

The P300 component of the ERP is the prototypical example of an endogenous component because its amplitude is modulated by subjective aspects of experimental stimuli, such as their salience, their task relevance, and their probability. The first P300 reported in the literature was a positivity that peaked 300 ms after stimulus onset. However, cognitive neuroscientists have subsequently observed a whole family of positive-going components of varying latency, all of which are referred to by the P300 moniker. P300s are elicited by any stimulus that requires the participant to make a binary decision. The amplitude of this response is proportional to the rarity of the target stimulus, as well as how confident the participant is in her classification judgment. The latency of the P300 (i.e. the point in time at which it peaks) varies with the difficulty of the categorization task, and ranges from 300 to over 1000 ms after the onset of the stimulus (Donchin, Ritter, & McCallum 1978; Kutas, McCarthy, & Donchin 1977; Magliero 1984; McCarthy 1981; Ritter et al. 1983).

While not specifically sensitive to language, P300s will be elicited in any psycholinguistic paradigm that requires a binary decision. As long as the experimenter is aware of the conditions known to modulate the P300, this component can serve as a useful dependent measure of language-relevant decision-making. For example, participants might be presented with a sentence, and then asked to judge its grammaticality. The amplitude of the P300 in such a case varies with the participant's confidence in her decision, and its latency indexes when the decision is made. This sort of experimental paradigm could thus be used to test hypotheses about the existence of a precisely characterized continuum of grammaticality. However, because P300 amplitude is very sensitive to stimulus probability, the number of critical stimuli in each experimental condition must be held constant. Another thing to be aware of is the fact that task-induced P300s may overlap in time with more specifically language sensitive ERP effects such as the N400.

Figure 3 shows ERPs elicited by the last word of sentences that ended either congruously or incongruously. Each graph shows the ERPs recorded at a different electrode site, as indicated in the little head. As is typical for these sorts of figures, the graph at the top of the figure corresponds to the electrode closest to the front of the head (where the nose is), and the graph at the bottom corresponds to the electrode over the back of the head. The solid line represents the average signal timelocked to the presentation of the last word in each congruous sentence, while the dotted line is the ERP to incongruous sentence completions. Visually presented words typically elicit 3 components, the N1, the P2 and the N400. The N1 and the P2 are thought to reflect different aspects of the visual processing of the words, and consequently are the same size for the congruous and the incongruous stimuli. The N400 is larger for the incongruous stimuli, and this component is thought to reflect the processing of the meaning of the stimuli.

3. Language sensitive ERPs

Since the discovery of the N400, cognitive neuroscientists interested in language have frequently appealed to ERPs as a dependent measure in psycholinguistic experiments. As a result, a number of language-sensitive ERP components have been reported (see (Kutas et al. 2000; Kutas & Van Petten 1994) for review). Although most of this research has been

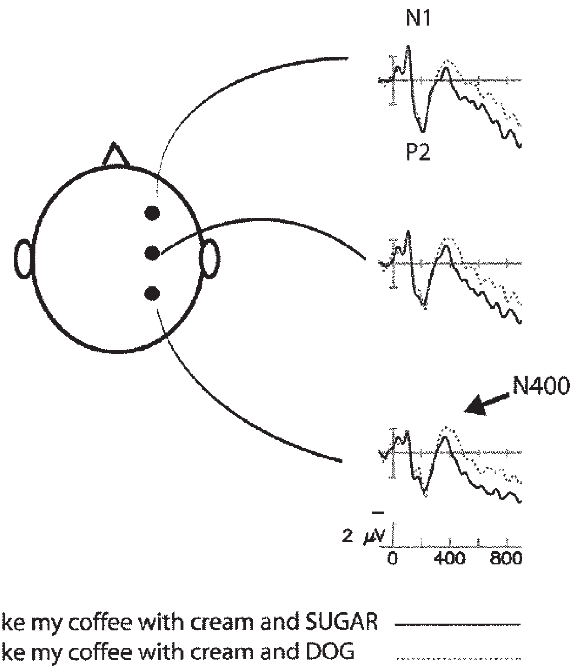


Figure 3. ERPs to the last word of congruous (solid) and incongruous (dotted) sentences

motivated by generative approaches to sentence processing, these finding may prove valuable to researchers interested in cognitive linguistics. Below we review ERP components known as the N400, the lexical processing negativity (LPN), the left anterior negativity (LAN), the P600, as well as slow cortical potentials, and briefly discuss the utility of each for studies of language comprehension.

3.1 N400

The N400 is a negative-going wave evident between 200 and 700 ms after the visual presentation of a word. Though this effect is observed all over the scalp, it is largest over centroparietal areas and is usually slightly larger on the right side of the head (Kutas, Van Petten, & Besson 1988). The N400 is elicited by words in all modalities, whether written, spoken, or signed (Holcomb & Neville 1990). Moreover, the size, or amplitude, of the N400 is affected in a way that is analogous in many respects to popular measures of priming in psycholinguistics, such as naming and lexical decision latencies.

For instance, in both word lists and in sentences, high frequency words elicit smaller N400s than low frequency words (Smith & Halgren 1989). The N400 also evidences semantic priming effects, in that the N400 to a word is smaller when it is preceded by a related word than when it is preceded by an unrelated word (Bentin 1987; Holcomb 1988). Third, the N400 is sensitive to repetition – smaller to subsequent occurrences of a word than to the first (Rugg 1985; C. Van Petten, Kutas, Kluender, Mitchiner, & McIsaac

1991). Further, while pseudowords (orthographically legal letter strings) elicit even larger N400s than do real words, orthographically illegal nonwords elicit no N400 at all (Kutas & Hillyard 1980).

In addition to its sensitivity to lexical factors, the N400 is sensitive to contextual factors related to meaning. For example, one of the best predictors of N400 amplitude for a word in a given sentence is that word's cloze probability (Kutas & Hillyard 1984). Cloze probability is the probability that a given word will be produced in a given context on a sentence completion task. The word "month" has a high cloze probability in "The bill was due at the end of the –," a low cloze probability in "The skater had trained for many years to achieve this –," and an intermediate cloze probability in "Because it was such an important exam, he studied for an entire –." N400 amplitudes are large for unexpected items, smaller for words of intermediate cloze probability, and are barely detectable for contextually congruous words with high cloze probabilities. In general, N400 amplitude varies inversely with the predictability of the target word in the preceding context.

Because initial reports of the N400 component involved the last word of a sentence, many people have the misconception that the N400 is an ERP component elicited by sentence final words. However, N400 is typically elicited by all words in a sentence. Interestingly, the size of the N400 declines across the course of a congruent sentence, starting large and becoming smaller with each additional open-class word. This effect has been interpreted as reflecting the buildup of contextual constraints as a sentence proceeds because it does not occur in grammatical but meaningless word strings (Van Petten & Kutas 1991). In general, the amplitude of the N400 can be used as an index of processing difficulty: the more demands a word poses on lexical integration processes, the larger the N400 component will be. This feature of the N400 makes it an excellent dependant measure in language comprehension experiments. As long as words in different conditions are controlled for length, frequency in the language, ordinal position in a sentence, and cloze probability, N400 amplitude can be used as a measure of processing effort. The N400 component is thus useful for testing hypotheses about the relative processing difficulty of minimal pairs of sentences.

Besides words, the N400 component is also elicited by pictorial stimuli. For example, line drawings of objects elicit a larger N400 when they are preceded by an unrelated word than by a related word (Praterelli 1994). Similarly, both line drawings and photographs of objects elicited a larger N400 when preceded by pictures of semantically unrelated than related objects (Holcomb & McPherson 1994; McPherson & Holcomb 1999). The N400 is larger for objects that occur in an incongruous scene (e.g. a desk in a river) than in a congruous one (e.g. a pot in a kitchen) (Ganis & Kutas 2003). Similarly, the N400 is larger for contextually inappropriate than appropriate objects in video clips of common activities, as in the contrast between a man shaving with a rolling pin versus a razor (Sitnikova, Kuperberg, & Holcomb 2003). Pictorial stimuli thus elicit the N400 component in a way that reflects the congruency of the pictured object with surrounding context, both linguistic and nonlinguistic.

3.2 LPN

The lexical processing negativity (LPN) is a brain potential to written words that is most evident over left anterior regions of the scalp. Its association with lexical processing derives from the fact that its latency is highly correlated with word frequency, peaking earlier for more frequent words (King & Kutas 1998). This component was originally thought to be an electrophysiological index of the brain's distinction between open-class content words and closed-class function words as the so-called N280 component was elicited by closed but not open class words (Neville, Mills, & Lawson 1992). However, subsequent testing indicated that word class effects are attributable to quantitative differences in word length and frequency (Osterhout, Bersick, & McKinnon 1997). That is, two words with the same frequency in the language elicit LPNs with the same latency even if one is an open-class word and the other a closed-class word (King & Kutas 1998). Because its latency is sensitive to word frequency, this component is useful as an indicator that the initial stages of lexical processing have been completed.

3.3 LAN

Researchers have also identified ERP components that seem to be sensitive to syntactic manipulations. The first is a negativity that occurs in approximately the same time window as the N400 (that is, 300–700 ms post-word onset) and is known as the LAN (left anterior negativity) because it is most evident over left frontal regions of the head. Kluender & Kutas described this component in a study of sentences containing long distance dependencies that required the maintenance of information in working memory during parsing (Kluender & Kutas 1993). Similarly, King & Kutas (King & Kutas 1995) described this component as being larger for words in object relative sentences like (3) that induce a greater working memory load than subject relative sentences like (4).

(3) The reporter who the senator attacked admitted the error.

(4) The reporter who attacked the senator admitted the error.

As an ERP component sensitive to working memory load, the LAN can be used to index differences in the processing difficulty of appropriately controlled stimuli.

3.4 P600

Another ERP component sensitive to syntactic and morphosyntactic processing is the P600. This slow positive shift has been elicited by violations of agreement, phrase structure, and subcategorization in English, German, and Dutch (Coulson, King, & Kutas 1998; Hagoort 1993; Mecklinger, Schriefers, Steinhauer, & Friederici 1995; Neville, Nicol, Barss, Forster, & Garrett 1991; Osterhout & Holcomb 1992). This component is typically described as beginning around 500 ms post-stimulus, and peaking at approximately 600 ms. Its scalp distribution tends to be posterior (larger over the back of the head), although anterior effects have also been reported (see Coulson et al. 1998 for review). Because the broad positivity is elicited by syntactic errors, it has been hypothesized to reflect a re-analysis of sentence structure triggered by such errors (Hahne & Friederici 1999).

However, Coulson and colleagues (Coulson et al. 1998) found that the amplitude of the P600 varied with the probability of ungrammatical trials within an experimental block. In one half of their study, 80% of the sentences were grammatical (the probable event), and only 20% ungrammatical (and improbable). In the other half of the study, 80% of the sentences were ungrammatical (and thus were the probable event), and only 20% grammatical (and thus improbable). The grammaticality effect (P600) was much smaller when ungrammatical sentences were probable than when they were improbable. In fact, the P600 to all improbable trials (collapsed across grammaticality) was very similar to the P600 to all ungrammatical trials (collapsed across probability).

Coulson and colleagues have argued that the P600 is not a syntax-specific component but rather a variant of a domain general component in the P300 family which has been hypothesized to reflect "context-updating," a process in which the subject recalibrates her expectations about the pattern of events in the environment (Donchin & Coles 1988). This interpretation is bolstered by recent evidence that P600 (but not N400) is elicited by semantic reversal anomalies such as "The cat that fled from the mice ran across the room," (Kolk, Chwilla, van Herten, & Oor 2003).

In any case, the fact that syntactic violations are associated with the late positive ERP known as the P600 provides a theory-neutral tool for testing when speakers find sentences to be ungrammatical, and potentially how manipulations of, for example, construal might have a graded effect on this electrophysiological index of ungrammaticality.

3.5 Slow cortical potentials

Besides phasic ERPs, temporally extended tasks such as reading or speech comprehension also elicit electrical changes with a slower time course. In order to examine slow brain potentials, it is necessary to average several seconds worth of data (viz. average EEG that begins at the onset of a particular class of language stimuli and ends several seconds later), apply a low-pass filter to the ERP data, and restrict analysis to activity less than 0.7 Herz. Kutas describes three slow brain potentials which have be elicited in experimental studies of written language comprehension (Kutas 1997). The first is a left lateralized negative shift evident at electrodes placed over occipital (visual) cortex, thought to reflect early visual processing. The second is the clause ending negativity (CEN), an asymmetric negativity larger over left hemisphere sites that may be associated with sentence wrap-up operations. The third is an ultra-slow (<0.2 Hertz) positivity over frontal sites that may be associated with discourse integration.

4. ERPs and cognitive linguistics

4.1 Frames

One important idea in cognitive linguistics is that words (and other sorts of linguistic structure) are not intrinsically meaningful, but are used by speakers to actively construct meaning. Lee (2002), for example, urges us to think of words as "tools that cause listeners to activate certain areas of their knowledge base, with different areas activated to different degrees in different contexts of use." Fillmore (Fillmore 1968) noted that understanding the meaning of many words requires an understanding of the concepts and conventions that surround their use. For example, "weekend" presumes an understanding of the structure of the week. Moreover, a true appreciation of the meaning of "weekend" involves cultural knowledge that many people in industrialized countries work Monday through Friday, but not on Saturday and Sunday.

In cognitive linguistics, frame semantics is a research program in which a word's semantic properties are described with respect to the way that they highlight aspects of an associated frame, or structured set of background assumptions. For example, "buy" and "sell" both evoke what might be dubbed the Commercial Transaction frame (Fillmore 1968). But "buy" highlights the buyer and the goods, while "sell" highlights the seller and the money. Langacker (Langacker 1987) argues that while "roe" and "caviar" refer to the same thing, their meanings differ because "roe" presumes a biological frame, while "caviar" presumes a culinary one. Conversely, Lakoff (Lakoff 1987)shows how many variants on the meaning of "mother," (as in "my birth mother" versus "my adopted mother") can arise because their meaning presumes different frames or idealized cognitive models of parenthood.

This tenet of cognitive linguistics thus suggests that background knowledge represented in frames figures prominently in the establishment of meaning, as language functions against this backdrop of conceptual structure. Although the psychological reality of frames has not been a big issue in language research, there is some ERP data consistent with the idea that background information represented in frames impacts language comprehension. In one such experiment, St. Georges, Mannes, & Hoffman (St. George, Mannes, & Hoffman 1994) recorded participants' ERPs as they read ambiguous paragraphs that either were or were not preceded by a disambiguating title. For example, people read paragraphs like the following (from (Bransford & Johnson 1972) "The procedure is actually quite simple. First you arrange items into groups. Of course one pile may be sufficient depending on how much there is to do.... After the procedure is complete one arranges the materials into different groups again. Then they can be put in their appropriate places. Eventually they will be used once more and the whole cycle will then have to be repeated...."

The title for this example is "Washing Clothes" and provides the subject with information about the appropriate frame to activate. Behavioral research by Bransford & Johnson showed that people both understood the paragraphs better and were better able to remember their contents when they were preceded by titles than when they were not. St. George and colleagues found that words in the untitled paragraphs elicited greater amplitude N400s than the same words in titled paragraphs. Although the local contextual clues provided by the paragraphs were identical in the titled and untitled conditions, ERPs revealed real-time processing differences in the lexical integration of words in the two conditions. These findings suggest that activating the frame facilitated the comprehension of these materials.

The study of joke comprehension also addresses the importance of frames for language comprehension, since many jokes are funny because they deceive the listener into using the wrong frame to help interpret the information presented in the first part of the joke. For example, consider the following joke, "I let my accountant do my taxes because it saves time: last spring it saved me ten years." Initially, the listener assumes a busy-professional frame is active; however, at the punch-line "saved me ten years" it becomes apparent that a crooked-businessman frame is more appropriate. Although lexical reinterpretation plays an important part in joke comprehension, to truly appreciate this joke it is necessary to recruit background knowledge about particular sorts of relationships that can obtain between business people and their accountants so that the initial busy professional interpretation can be mapped into the "crooked-businessman" frame. Coulson refers to the semantic and pragmatic reanalysis needed to understand examples like this one as frame-shifting (Coulson 2001).

Given the impact of frame-shifting on the interpretation of one-line jokes, one might expect the underlying processes to take time, and consequently be reflected in increased reading times in behavioral tests of processing difficulty such as self-paced reading. In this paradigm, the task is to read sentences one word at a time by pressing a button to advance to the next word. As each word appears, the preceding word disappears, so that the experimenter gets a record of how long the participant spent reading each word in the sentence. Coulson & Kutas (1998) used this method to compare reading times for sentences that ended as jokes with reading times for the same sentences with non-funny "straight" endings that were consistent with the contextually evoked frame. Two types of jokes were tested, high constraint jokes like (5) which elicited at least one response on a sentence completion task with a cloze probability of greater than 40%, and low constraint jokes like (6) which elicited responses with cloze probabilities of less than 40%. For both (5) and (6) the word in parentheses is the most popular response on the cloze task.

(5) I asked the woman at the party if she remembered me from last year and she said she never forgets a (face 81%).

(6) My husband took the money we were saving to buy a new car and blew it all at the (casino 18%).

To control for the fact that the joke endings are (by definition) unexpected, the straight controls were chosen so that they matched the joke endings for cloze probability, but were consistent with the frame evoked by the context. For example, the straight ending for (5) was name (the joke ending was *dress*); while the straight ending for (6) was tables (the joke ending was *movies*). The cloze probability of all four ending types (high and low constraint joke and straight endings) was equal, and ranged from 0%–5%. Coulson & Kutas (Coulson & Kutas 1998) found that readers spent longer on the joke than the straight endings, and that this difference in reading times was larger and more robust in the high constraint sentences. This finding suggests there was a processing cost associated with frame-shifting reflected in increased reading times for the joke endings, especially in high constraint sentences that allow readers to commit to a particular interpretation of the sentence.

In a very similar ERP study of the brain response to jokes, Coulson & Kutas (Coulson & Kutas 2001) found that ERPs to joke endings differed in several respects from those to the straight endings, depending on contextual constraint as well as participants' ability to get the jokes. In poor joke comprehenders, jokes elicited a negativity in the ERPs between 300 and 700 milliseconds after the onset of the sentence-final word. In good joke comprehenders, high but not low constraint joke endings elicited a larger N400 than the straight endings. Also, in this group, both sorts of jokes (high and low constraint) elicited a late positivity in the ERP (500–900 ms post-onset) as well as a slow sustained negativity over left frontal sites. Multiple ERP effects of frame-shifting suggest the processing difficulty associated with joke comprehension involves multiple neural generators operating with slightly different time-courses.

Taken together, these studies of frame-shifting in jokes are far more informative than either study alone. The self-paced reading time studies suggested that frame-shifting needed for joke comprehension exerts a processing cost that was especially evident in high constraint sentence contexts (Coulson & Kutas 1998). ERP results suggested the processing cost associated with frame-shifting is related to higher-level processing (Coulson & Kutas 2001). In the case of the high constraint jokes, the difficulty includes the lexical integration processes indexed by the N400, as well as the processes indexed by the late-developing ERP effects. In the case of the low constraint jokes, the difficulty was confined to the processes indexed by late-developing ERP effects. The added difference in lexical integration indexed by the N400 may explain why joke effects on reading times were more pronounced for high constraint sentences than for low. Because the late developing ERP effects were only evident for good joke comprehenders who successfully frame-shifted, they are more likely to be direct indices of the semantic and pragmatic reanalysis processes involved in joke comprehension.

As a general methodological point, the demonstration of individual differences in memory, vocabulary, language ability, or, in this case, on-line comprehension, can be extremely informative when they can be correlated with particular ERP effects. Coulson & Kutas (Coulson & Kutas 2001), for example, were able to determine that the N400 joke effect was not an index of successful frame-shifting as it was evident in ERPs collected from both good and poor joke comprehenders. In contrast, the late positivity present only in good comprehenders' ERPs was argued to be more closely related to the frame-shifting process important for understanding jokes. Researchers in experimental cognitive linguistics would do well to consider how individual differences in cognitive abilities affect various language comprehension phenomena and their manifestations in the ERP signal. Moreover, work on joke comprehension by Coulson & Kutas demonstrates how ERPs and reaction time data for the same stimuli can provide complementary information about the underlying cognitive processes.

4.2 Metaphor

In fact, ERPs can reveal reliable differences even when no reaction time differences are evident. This is important because reaction times are typically interpreted as reflecting processing difficulty, yet it is quite possible for two processes to take the same amount of time, but for one to recruit more neural processing resources. One issue in cognitive linguistics where this has been an important issue is the study of metaphor comprehension. Because classical accounts of metaphor comprehension (Grice 1975; Searle 1979) depict a two-stage model in which literal processing is followed by metaphorical processing, many empirical studies have compared reading times for literal and non-literal utterances and found that when the metaphorical meaning was contextually supported, reading times were roughly similar. However, as Gibbs notes, parity in reading times need not entail parity in the underlying comprehension processes (Gibbs 1994). It is possible, for example, that literal and metaphorical meaning might take the same amount of time to comprehend, but that the latter required more effort or processing resources. Alternately, comprehension processes for literal versus metaphorical utterances might take the same amount of time to complete, and yet involve quite different computations (Gibbs & Gerrig 1989).

Because they involve a direct and continuous measure of brain activity, ERPs can potentially distinguish between qualitatively different sorts of processing, even if their corresponding behavioral manifestations require the same amount of time. Taking advantage of the known relationship between N400 amplitude and processing difficulty, Pynte and colleagues contrasted ERPs to familiar and unfamiliar metaphors in relevant versus irrelevant contexts (Pynte et al. 1996). They found that regardless of the familiarity of the metaphors, N400 amplitude was a function of the relevance of the context. Moreover, by using ERPs, Pynte and colleagues employed a measure which is in principle capable of revealing the qualitative processing differences by the standard (Gricean) pragmatic model. In fact, they observed no evidence of a qualitative difference in brain activity associated with the comprehension of literal and metaphoric language.

Reports that literal and metaphorical language comprehension display a similar time course and recruit a similar set of neural generators are consistent with a number of modern models of metaphor comprehension (Coulson & Matlock 2001; Gibbs 1994; Giora 1997; Glucksberg 1998). Coulson's (Coulson 2001) model also makes predictions for comprehension difficulty, predicting a gradient of processing difficulty related to the extent to which comprehension requires the participant to align and integrate conceptual structure from different domains. This prediction was tested by Coulson & Van Petten (Coulson & Van Petten 2002) when they compared ERPs elicited by words in three different sentence contexts on a continuum from literal to figurative, as suggested by conceptual blending theory (Fauconnier & Turner 2002). For the literal end of the continuum, Coulson & Van Petten used sentences that promoted a literal reading of the last term, as in "He knows that whiskey is a strong INTOXICANT." At the metaphoric end of the continuum, they used sentences which promoted a metaphoric reading of the last term, as in "He knows that power is a strong INTOXICANT." Coulson & Van Petten also posited a literal mapping condition hypothesized to fall somewhere between the literal and the metaphoric uses, such as, "He has used cough syrup as an INTOXICANT."

Literal mapping stimuli employed fully literal uses of words in ways that were hypothesized to include some of the same conceptual operations as in metaphor comprehension. These sentences described cases where on object was substituted for another, one object was mistaken for one another, or one object was used to represent another – all contexts that require the comprehender to set up mappings between the two objects in question, and the domains in which they typically occur. In positing a continuum from literal to metaphorical based on the difficulty of the conceptual integration needed to comprehend the statement, Coulson & Van Petten (2002) predicted a graded difference in N400 amplitude for the three sorts of stimuli.

Data reported by Coulson & Van Petten were largely consistent with these predictions. In the early time window, 300–500 ms post-onset and before, ERPs in all three conditions were qualitatively similar, displaying similar waveshapes and scalp topography. This suggests that during the initial stages, processing was similar for all three sorts of contexts. Moreover, as predicted N400 amplitude differed as a function of metaphoricity, with literals eliciting the least N400, literal mappings the next-most, and metaphors eliciting the most N400, suggesting a concomitant gradient of processing difficulty. The graded N400 difference argues against the literal/figurative dichotomy inherent in the standard model, and is consistent with the suggestion that processing difficulty associated with figurative language is related to the complexity of mapping and conceptual integration.

4.3 Iconicity

One way that cognitive linguistics differs from more traditional approaches to language is the assumption that the relationship between linguistic forms and their meaning is often iconic. An iconic relationship between a symbol and its referent is one in which the symbol shares some structure with the thing that it represents. While linguists have traditionally assumed the relationship between words and their referents is arbitrary and established by convention, cognitive linguists posit a much higher degree of iconicity in language. It is assumed that the presence of iconic motivation for linguistic forms makes it easier to learn a language, and provides a basis for language change over time.

Langacker (Langacker 1997, 1999) argues that because all conceptualization unfolds dynamically in real time, the temporal dimension of language is a critical one that is bound to have linguistic ramifications. Langacker (Langacker 2003) notes that one of the most obvious of these is the iconic use of word order to suggest the sequence of events. For example, contrast (7) and (8).

(7) She got married and had a baby.

(8) She had a baby and got married.

Of course, Langacker does not suggest that linguistic constructions must be iconic in this way. Rather, he claims that because the function of linguistic constructions is to guide conceptualization, the dynamic nature of conceptualization imposes certain constraints on linguistic structure. For instance, Langacker (Langacker 2003) argues that (9a) is more natural than (9b), because in the former the meaning of "from" and "to" serve to reinforce the iconically cued pattern of mental scanning (knee → ankle), whereas they conflict with it in (9b).

(9) a. A scar extends from his knee to his ankle.
b. A scar extends to his knee from his ankle.

While these claims are perfectly plausible, they rest on the assumption that certain forms of conceptualization are more natural than others, and predict differences in processing difficulty as a function of linguistic factors. Langacker's claims, then, are exactly the sort of ideas that can and should be tested with empirical methods such as ERP experiments.

Indeed, Münte, Schiltz, & Kutas (1998) have used slow cortical potentials evident in recorded ERPs to address whether temporal iconicity affects the processing of language. Münte and colleagues hypothesized that people's conception of time as a sequential order of events determines the way we process statements referring to the temporal order of events. Consequently, they recorded ERPs as participants read sentences such as "Before/After the psychologist submitted the article, the journal changed its policy." Because, "Before X,Y" presents information in the reverse chronological order, it was hypothesized that these sentences would be more difficult to process than the "After" sentences. Indeed, Münte et al. found that ERPs recorded at electrode sites on the left frontal scalp were more negative for the more difficult "Before" sentences. Perhaps more compelling, they found that the size of this effect was correlated with individual participants' working memory spans. As noted above, individual differences that correlate with ERP effects may illuminate the functional significance of those effects in a way that bears on the original question. In this case, the data suggest that temporally iconic "after" sentences imposed less of a demand on working memory than did the less iconic "before" ones.

5. Considerations on ERP research

Although the previous section's review of cognitive linguistic ERP research was not necessarily exhaustive, it was fairly comprehensive. While intriguing, this handful of studies hardly constitutes a complete body of research. Although there are known limitations to using ERP data to localize neural generators in the brain, it is an excellent measure for determining precisely when the processing of two classes of stimuli begins to diverge. Because it can provide a continuous real time index of processing that occurs at the advent of a linguistic stimulus, the ERP is well-suited for addressing questions that have to do with what sorts of information experimental participants are sensitive to and when.

Because the N400 is sensitive to the same processes indirectly assessed in the reaction time paradigm, we can view its use in investigations of language comprehension as an analogous version of behavioral measures such as priming. One advantage of ERP measures, however, is that they can be collected in the absence of an explicit task (other than that of language comprehension itself). Moreover, ERP measures can also be collected while the participant performs a behavioral task, thus giving the experimenter a measure of on-going brain activity before, during, and after the performance of the task. Regardless of whether one conducts two experiments – one behavioral and one ERP – or, whether the two sorts of measures are collected together, ERP data can greatly aid in the interpretation of the behavioral results.

In fact, ERP and reaction time data are often complementary as reaction time data can provide an estimate of how long a given processing event took, while ERP data can suggest whether distinct processes were used in its generation. An experimental manipulation that produces a reaction time effect might produce two or more ERP effects, each of which is affected by different sorts of manipulations. By giving the experimenter the means to explore these dissociations, ERPs can help reveal the cognitive processes that underlie the linguistic phenomena of interest. In fact, because they can be identified with specific cognitive processes (e.g., meaning integration, working memory load, or the registration of an ungrammatical sentence), ERP effects provide evidence of how processing differs in the experimental conditions (King & Kutas 1995).

5.1 Constraints on ERP research

The basic ERP methodology in language research involves recording ERPs to minimally different sorts of language stimuli in order to observe modulations in the amplitude and/or latency of particular components. However, there are a number of constraints to keep in mind when designing an ERP experiment. First, the subject must remain relatively still during the recording of the EEG. Because the signal in the ERPs is very small (typically less than 10 microvolts), the bioamplifiers are extremely sensitive. Movement of the body can result in temporarily overloading the amplifiers.

Further, because the eye also has electrical properties, blinking and eye movements produce electrical activity that is detected by the EEG electrodes. While the brain is encased in the skull, the eyes are not. Consequently, voltage fluctuations due to eye movements can obscure the much more theoretically interesting brain activity. Although there are a few different mathematical techniques for isolating the contribution of the eyes to the EEG data, most ERP researchers deal with this problem by telling participants not to blink or move their eyes during critical points in the experiment, and by not including EEG corrupted by blinks or eye movements in the ERP.

The larynx also has electrical properties detectable by EEG electrodes so that it is not practical to record ERPs to the production of overt speech. What's more speech production can result in body motion that results in electrical noise and overloading the bioamplifiers. It is possible, however, to record ERPs to the onset of a stimulus for naming, such as a word or a picture. If naming latencies are short, the ERP elicited by the picture is likely to be corrupted by noise induced by speech onset. However, slightly longer naming latencies enable the researcher to collect artifact-free ERP. Although the ERP technique is more compatible with studying language comprehension, a number of researchers have developed paradigms for addressing issues in language production (see e.g. Jescheniak, Schriefers, Garrett, & Friederici 2002; Schmitt, Schlitz, Zaake, Kutas, & Münte 2001).

Another constraint to keep in mind is that – by definition – the ERP is the brain response to numerous stimuli that share some theoretically interesting property such as occurring in a true sentence rather than a false one, or being a prototypical category member as opposed to a non-prototypical one. For language experiments, a minimum of 30 trials in each experimental condition (that is, each "cell" in the design) is recommended to obtain a reasonable ratio of signal to noise. As several components of the ERP are sensitive to stimulus repetition (VanPetten 1991), most experimenters construct multiple stimulus lists in order to fully counterbalance their design without requiring individual subjects to read multiple versions of a single stimulus. Also, because ERPs can vary greatly between individuals, it is advisable to use a within-subjects design whenever possible. That is, all subjects should participate in (viz. read examples from) every condition so as to minimize the impact of individual differences on the experimental results.

Because ERPs require a discrete processing event for their elicitation, it is best if experimental effects are predicted to occur on a particular word or words in the sentence. In experiments where participants read the stimuli, sentences are usually presented one word at a time so that the ERP is time-locked to the critical word in the sentence. In experiments using spoken materials, it is important for the critical word or words to have a clear onset.

Recording ERPs to auditorally presented materials comes with its own set of challenges. Perhaps the main problem is that words in continuous speech do not generally elicit distinct ERPs because word boundaries are often absent from the speech signal. Fortunately, it is still possible to observe measurable differences in N400 amplitude to the last word of congruous and incongruous sentence completions (e.g. Holcomb & Neville 1990; Van Petten, Coulson, Rubin, Plante, & Parks 1999). Moreover, Müller and colleagues point to the utility of examining slow cortical potentials when investigating the comprehension of spoken language (Müller, King, & Kutas 1997). In a sentence processing study that compared ERPs elicited by subject-relative sentences with the more demanding object-relative sentences, Müller and colleagues identified an ultra-slow frontal positivity whose amplitude varied as a function of comprehension difficulty in both written and spoken materials.

Similarly, Steinhauer and colleagues (Steinhauer, Alter, & Friederici 1999) have used ERPs to study how intonational phrasing guides the initial analysis of sentence structure. They recorded ERPs as subjects listened to syntactically ambiguous sentences with appropriate and inappropriate prosodic cues. In naturalistic stimuli, Steinhauer and colleagues found that participants' ERPs showed a positive-going waveform at prosodic boundaries that they call the Closure Positive Shift. In cases where the prosodic cues conflicted with syntactic ones, the mismatch elicited an N400-P600 pattern of ERP components suggesting participants used prosodic features to determine their initial (incorrect) parse of the sentence. These results show that the ERP is a good measure for revealing the time course and neural basis of prosodic information. These factors could prove useful for the study of many topics in cognitive linguistics.

5.2 Future directions

One promising area for future ERP research is to test hypotheses stemming from the central tenet of cognitive linguistics that linguistic forms are often motivated by particular contrasts in meaning. For instance, though the distinction between count nouns (like cat and shoe) and mass nouns (like water and footwear) is arbitrary to an extent, it seems to be motivated by the way in which their referents are construed. While count nouns are typically used to denote unitary objects that can be individuated, mass nouns often denote homogeneous substances that are difficult or impossible to individuate. Interestingly, non-prototypical uses of mass nouns seem to support the idea that the count/mass noun distinction marks a difference in construal of the target object (Langacker 1987). For instance, uncooked potatoes are easy to individuate and function as count nouns. Mashed potatoes, however, form a homogeneous substance and can be referred to with a mass noun, as in (10).

(10) Could I please have some more potato?

The semantic motivation for the mass/count distinction might be tested by combining drawings or photographs that promote a particular construal of the target object with language stimuli that employ either a mass or a count noun. If the construal of the target object is relevant to the choice of a mass versus a count noun, any incompatibility between the construal promoted by the pictures and that promoted by the linguistic form ought to give rise to an ERP effect. One might present the pictures first and time-lock ERPs to the onset of a critical word in the sentences that employ mass vs. count nouns. Alternatively, one might present the language stimulus first, and time-lock ERPs to the onset of pictures that are either compatible or incompatible with the construal promoted by the preceding sentence.

The proposed combination of language and pictures would seem to be quite promising for cognitive linguistic research. Basic proposals about the importance of perspective and profiling in meanings constructed by speakers could be tested by contrasting ERPs to construal-appropriate and construal-inappropriate pictures. Similarly, photographs could be used to promote the activation of background knowledge from various domains to test the way in which it affects the interpretation of words with metaphoric, metonymic, or otherwise polysemic meanings.

Although cognitive neuroscientists have learned a great deal about language comprehension, it remains the case that most studies have employed experimenter-constructed stimuli in the controlled and artificial setting of the laboratory. Naturalistic language use can differ markedly from that employed in most scientific studies. Not only are the units of processing larger than those typically studied – that is, texts and discourses rather than the words and sentences so dear to the hearts of psycholinguists – but there are social and physical cues to guide the language user. In the future, we must exploit technological advances to bring more of the world into the laboratory. For instance, using mpeg (Moving Picture Experts Group audio Layer-3) technology it is possible to present subjects with auditory stimuli, such as naturally occurring conversation, or more controlled, scripted versions of the same phenomena.

Another facet of normal language comprehension typically absent from laboratory studies is the presence of visual information. This visual information includes both the local context as well as visual information about the speaker, such as her facial expressions and her gestures. As EEG can in principle be time-locked to the onset of visual events in MP3 videos, it is possible to record ERPs as subjects watch videos of speakers interacting in real contexts. Although the continuous nature of videographic stimuli present some of the same problems as continuous speech, it seems plausible that large differences in processing difficulty would be evident in ERP effect to visual stimuli, just as they are to speech.

Though we have seen that the 10% myth is false, and that the brain is ultimately an organ whose cells engage in a continual exchange of electrochemical signals, there is some truth to the idea that our brains harbor unharnessed potential for learning about all sorts of things – including brain function itself. As discussed in this chapter, electromagnetic changes in the brain can be detected with electrodes placed on the scalp. Remarkably, the ERP technique is sensitive to the cognitive operations involved in language comprehension and suitable for addressing various issues in cognitive linguistics. In particular, the on-going nature of the EEG signal makes it a good dependent measure for assessing the comprehension of linguistic materials in real time. We have discussed some of the ways in which ERPs have already been used to test ideas in cognitive linguistics, as well as some ways ERPs might potentially be used to do so in the future. Clearly, the use of ERPs in this way has not reached even 10% of its potential! In the end, investigation of these issues is limited only by the imagination of the experimenters.

References

Bentin, S. (1987). Event-related potentials, semantic processes, and expectancy factors in word recognition. *Brain & Language, 31*, 308–327.

Bransford, J. D. & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior, 11*, 717–726.

Coulson, S. (2001). *Semantic Leaps: Frame-shifting and Conceptual Blending in Meaning Construction*. Cambridge, UK: Cambridge University Press.

Coulson, S., King, J., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language & Cognitive Processes, 13*(1), 21–58.

Coulson, S. & Kutas, M. (1998). *Frame-shifting and Sentential Integration* (Cognitive Science Technical Report No. 98.02). La Jolla, CA: UCSD.

Coulson, S. & Kutas, M. (2001). Getting it: Human event-related brain response in good and poor comprehenders. *Neuroscience Letters, 316*, 71–74.

Coulson, S. & Matlock, T. (2001). Metaphor and the space structuring model. *Metaphor & Symbol, 16*, 295–316.

Coulson, S. & Van Petten, C. (2002). Conceptual integration and metaphor: An ERP study. *Memory & Cognition, 30*(6), 958–968.

Donchin, E. & Coles, M. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*, 357–374.

Donchin, E., Ritter, W., & McCallum, C. (1978). Cognitive psychophysiology: The endogenous components of the ERP. In P. T. E. Callaway & S. H. Koslow (Ed.), *Brain event-related potentials in man*. New York: Academic.

Fauconnier, G. & Turner, M. (2002). *The Way We Think*. New York: Basic Books.

Fillmore, C. J. (1968). The case for case. In E. B. R. T. Harm (Ed.), *Universals of Linguistic Theory* (pp. 1–90). New York: Holt, Rinehart, and Winston.

Ganis, G. & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Brain Research: Cognitive Brain Research, 16*, 123–144.

Gibbs, R. (1994). *The Poetics of Mind: Figurative Thought, Language, and Understanding*. Cambridge, UK: Cambridge University Press.

Gibbs, R. & Gerrig, R. (1989). How context makes metaphor comprehension seem "special". *Metaphor & Symbolic Activity, 4*(3), 145–158.

Giora, R. (1997). Understanding figurative language: The graded salience hypothesis. *Cognitive Linguistics, 7*(1), 183–206.

Glucksberg, S. (1998). Understanding metaphors. *Current Directions in Psychological Science, 7*, 39–43.

Grice, H. (1975). Logic and conversation. In P. C. J. Morgan (Ed.), *Syntax and Semantics: Vol. 3, Speech Acts*. New York: Academic Press.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*(4), 439–483.

Hahne, A. & Friederici, A. (1999). Electrophysiological Evidence for Two Steps in Syntactic Analysis: Early Automatic and Late Controlled Processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205.

Hillyard, S. & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology, 34*, 33–61.

Holcomb, P. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain & Language, 35*, 66–85.

Holcomb, P. & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain & Cognition, 24*, 259–276.

Holcomb, P. & Neville, H. (1990). Semantic priming in visual and auditory lexical processing. *Language and Cognitive Processes, 5*, 281–312.

Jescheniak, J. D., Schriefers, H., Garrett, M. F., & Friederici, A. D. (2002). Exploring the activation of semantic and phonological codes during speech planning with event-related brain potentials. *Journal of Cognitive Neuroscience, 14*, 951–964.

King, J. & Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376–395.

King, J. & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters, 244*, 61–64.

Kluender, R. & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependancies. *Journal of Cognitive Neuroscience, 5*(2), 196–214.

Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Structure and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*, 1–36.

Kutas, M. (1997). Views on how the electrical activity that the brain generates reflects the functions of different language structures. *Psychophysiology, 34*(4), 383–398.

Kutas, M., Federmeier, K., Coulson, S., King, J. W., & Muente, T. F. (2000). Language. In G. G. Berntson (Ed.), *Handbook of Psychophysiology* (2nd ed., pp. 576–601). Cambridge, UK: Cambridge University Press.

Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*, 792–797.

Kutas, M. & Van Petten, C. (1994). Psycholinguistics electrified. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*, 218–233.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

Langacker, R. W. (1997). A dynamic account of grammatical function. In J. Bybee, J. Haiman, & S. A. Thompson (Eds.), *Essays on Language Function and Language Type Dedicated to T. Givon* (pp. 191–222). Amsterdam & Philadelphia: John Benjamins.

Langacker, R. W. (1999). *Grammar and Conceptualization*. Berlin and New York: Mouton de Gruyter.

Langacker, R. W. (2003). Dynamicity, fictivity, and scanning: The imaginative bases of logic and linguistic meaning. *Korean Linguistics, 18*, 1–64.

Lee, D. (2002). *Cognitive Linguistics: An Introduction*. Oxford: Oxford University Press.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*, 171–186.

McCarthy, G. D., E. (1981). A metric of thought: A comparison of P300 latency and reaction time. *Science, 211*, 77–80.

McPherson, W. B. & Holcomb, P. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*, 53–65.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition, 23*, 477–494.

Müller, H., King, J., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Brain Research: Cognitive Brain Research, 5*(3), 192–203.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature, 395*(6697), 71–73.

Neville, H., Mills, D., & Lawson, D. (1992). Fractionating language: Different neural subsystems with different sensitive periods. *Cerebral Cortex, 2*(3), 244–258.

Neville, H., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology, 46*(2), 143–168.

Osterhout, L. & Holcomb, P. (1992). Event-related potentials and syntactic anomaly. *Journal of Memory and Language, 31*, 785–804.

Praterelli, M. E. (1994). Semantic processing of pictures and spoken words: Evidence from event-related brain potentials. *Brain & Cognition, 24*, 137–157.

Pynte, J., Besson, M., Robichon, F., & Poli, J. (1996). The time-course of metaphor comprehension: An event-related potential study. *Brain & Language, 55*, 293–316.

Ritter, W., Simpson, R., & Vaughan, H. G. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology, 20*, 168–179.

Rugg, M. D. (1985). The effects of semantic priming and word repetition on event-related potentials. *Psychophysiology, 22*, 642–647.

Rugg, M. D. & Coles, M. (Eds.). (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford, UK: Oxford University Press.

Schmitt, B. M., Schlitz, K., Zaake, W., Kutas, M., & Münte, T. F. (2001). An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience, 13*, 510–522.

Searle, J. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge, UK: Cambridge University Press.

Sitnikova, T., Kuperberg, G., & Holcomb, P. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*, 160–164.

Smith, M. E. & Halgren, E. (1989). Dissociation of recognition memory components following temporal lobe lesions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 50–60.

St. George, M., Mannes, S., & Hoffman, J. (1994). Global semantic expectancy and language comprehension. *Journal of Cognitive Neuroscience, 6*, 70–83.

Steinhauer, K., Alter, K., & Friederici, A. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience, 2*(2), 438–453.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*(2), 394–417.

Van Petten, C. & Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory & Cognition, 19*, 95–112.

Van Petten, C., Kutas, M., Kluender, R., Mitchiner, M., & McIsaac, H. (1991). Fractionating the word repetition effect with event-related potentials. *Journal of Cognitive Neuroscience, 3*(2), 131–150.

VanPetten, C., Kutas, M. (1991). Influences of semantic and syntactic context on open and closed class words. *Memory & Cognition, 19*, 95–112.

Bridging language with the rest of cognition

Computational, algorithmic and neurobiological issues and methods

Shimon Edelman

1. Introduction

The possibility of integrating linguistics into a unified science of cognition – a desideratum put forward in many of the relevant disciplines – depends on the degree to which common computational principles (Marr & Poggio 1977) and brain mechanisms are shared by language and by the other cognitive functions. To explore this possibility, we need to bring together ideas from several fields, which as yet have seen little intellectual cross-fertilization. The first of these is cognitive linguistics (Langacker 1987; Bernárdez 1999) – a natural home discipline for the integration project, which consistently produces valuable insights into the psychology of language, yet is little concerned with algorithmic or implementational issues. The second is computational linguistics (Jurafsky & Martin 2000), including statistical natural language processing (Manning & Schütze 1999) – a field that examines the mathematical nature of language-related tasks and generates important applications, yet pays little attention to behavioral or neurobiological issues. Lastly, there is the Marr-Poggio computational framework (Marr & Poggio 1977), which is used across cognition and which spans all the relevant levels of analysis, but has not yet been extended to the study of language.

This chapter discusses some of the general computational principles that emerge as useful for understanding cognition, focusing on those that are likely to be especially relevant in dealing with structured knowledge. It then brings these principles to bear on a theory of language that is rooted both in cognitive and in computational linguistics, and that views language as an incrementally learnable system of redundant, distributed representations akin to those found by neurobiologists in olfaction, audition, vision, and motor control.

2. Common principles of cognitive representation and processing

The view that cognition hinges on the representation of knowledge by the brain is widely accepted in linguistics, psychology, neuroscience, and the philosophy of mind (Chomsky 1957; Miller 1962; Shepard 1975; Marr 1982; Gallistel 1990; Cummins 1996). Most importantly, representations play a central role in those theories of mind/brain that construe cognition as computation defined over representational states (Baum 2004).[1] A representational state in a cognitive system is characterized by its covariation with certain aspects of the relevant state of affairs in the world, and, crucially, by having counterfactually supported observable effects.[2]

2.1 How to garner empirical support for posited representations

Whereas thirty years ago linguists were expected to prove that the representations they posit are psychologically real (Fodor et al. 1974) by predicting and then demonstrating such effects, contemporary formal linguistics has, lamentably, given up on this requirement (Edelman & Christiansen 2003). Consider, for example, the following passage from an online introduction to a course in neurolinguistics: "We know already what isn't the right question: What is the psychological reality of linguistic entities and operations?"[3] As a result, stipulated linguistic *entia*, from Deep Structure and transformations in the 1960s to Logical Form, traces, Move, and Merge in the 1990s, have multiplied over the decades, arguably *praeter necessitatem*.[4] To date, none of the rare attempts to obtain psycholinguistic (behavioral) evidence for such entities have yielded unequivocal results. For example, in the recent study by (Nakano et al. 2002), only 24 subjects out of the original 80 performed consistently with the predictions of a trace/movement theory, while 39 subjects exhibited the opposite behavior (the data from the rest of the subjects were discarded because their error rate was too high).

The shakiness of the empirical foundations of generative linguistics appears to be especially disappointing when seen in the broader context of successful representational

1. Such states need not, and probably cannot, be wholly internal to the brain; cf. "The primary function of perception is to keep our internal framework in good registration with that vast external memory, the external environment itself" (Reitman et al. 1978: 72). Thus, in many respects, the world is its own best representation (O'Regan 1992).

2. A counterfactual is a logical conditional statement whose antecedent is taken to be contrary to fact by those who utter it; cf. "If linguistics were what the author claims, syntactic trees would be visible in CAT scans." (Postal 2004: Ch. 11).

3. Quote found at a web site for *Neurolinguistics*, course #24.944, taught by Alec Marantz (Head, Department of Linguistics and Philosophy, MIT) in 2000; see http://web.mit.edu/linguistics/www/marantz/marantz.home.24944f00intro.html.

4. Occam's Razor, often stated as *entia non sunt multiplicanda praeter necessitatem* [entities should not be multiplied beyond necessity], is a fundamental principle of the scientific method, which has recently assumed a central role in statistical inference and learning theory (Rissanen, 1987; Blumer et al., 1987).

theories that have emerged in other cognitive domains. Indeed, the need to demonstrate the psychological (behavioral), and, eventually, the neurobiological, reality of the theoretical constructs exists in all of cognition, including human vision, where, as in language, direct observation of the underlying mechanisms is difficult. An excellent example of how vision scientists have risen to this challenge is found in the history of the concept of multiple parallel spatial frequency channels (Figure 1), a representational hypothesis that had been introduced in the late 1960s, then completely vindicated by purely behavioral means over the following decade; see, e.g., (Wilson & Bergen 1979).

More generally, the logic of looking for empirical signatures of the posited representations can be put to work in a number of ways that are all well known to cognitive scientists. For example, the reality of a distinction between two representations or processes can be indicated by a *double dissociation*, that is, a situation in which each of the two can be obtained in isolation from the other, either as a result of "complementary" lesions in different patients (Damasio & Tranel 1993), or through experimental manipulation of stimuli presented to normal subjects (Pulvermüller et al. 1996). Likewise, one can use *priming* (Tulving & Schacter 1990; Ochsner et al. 1994): if a representation can be primed – that is, if its manifestation in response to a stimulus can be modified by prior exposure to a related stimulus – then it is real, and can be accounted for by the mechanism that embodies the memory trace for this class of stimuli (Wiggs & Martin 1998). Finally, the worries of some cognitive scientists that representations are merely epiphenomenal to cognition can be assuaged ultimately by demonstrating the *causal effectiveness* of representational mechanisms through direct intervention, such as microampere-level current injection at the appropriate brain site which brings about the predicted perceptual/behavioral change (Salzman et al. 1990).

2.2 Some common characteristics of cognitive representations

What kinds of representations does one find in the brain? In domains as diverse as olfaction, vision, reasoning and memory, the representations are typically *distributed* in that an ensemble of neurons (rather than a single neuron) is involved in each task, *overlapping* in that the response profiles of the members of an ensemble are redundant (rather than mutually exclusive), and *graded* in that each neuron's response depends smoothly (rather than in an all or none fashion) on the represented quantity. Thus, theories involving distributed, overlapping, and graded representations (Pouget et al. 2000) have enjoyed the most consistent and wide ranging explanatory success across cognition.

While thinking about distributed representations, it is important to realize that embracing the terminology of parallel distributed processing, often referred to as "connectionism", does not by itself constitute a particularly illuminating explanation. That the brain is, on the level of mechanism, a connectionist device is a trivial observation; understanding it will take coordinated action on computational (problem) and algorithmic (process) levels as well (Marr & Poggio 1977). This is precisely what is happening now in the cognitive sciences: researchers are converging on a few classes of problems to which various aspects of cognition can be reduced, and are forging an understanding of a few classes of computational processes, operating over distributed representations, that are

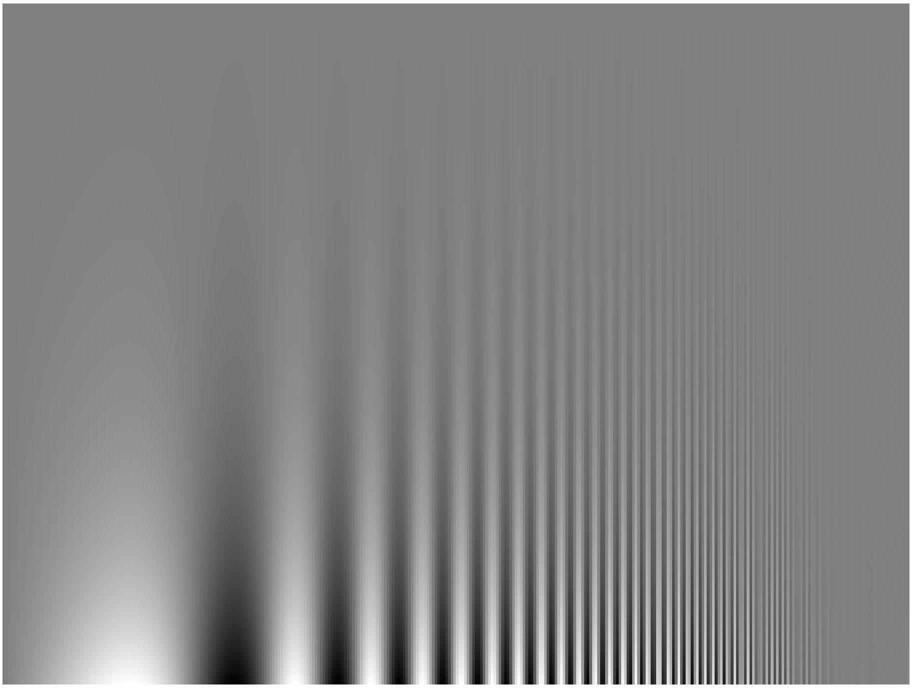


Figure 1. An illustration of the basic concepts required for a behavioral demonstration of the psychological reality of spatial frequency channels, which is a kind of visual representation found in the brain. In the image shown here, spatial frequency (rate of change of intensity across space) varies along the abscissa and contrast (difference between dark and light) along the ordinate (Campbell & Robson 1968). As you can see, more contrast is required to perceive the grating (the alternation between dark and light) for low and high frequencies, compared to intermediate frequencies. Researchers have postulated early on that the perception of spatial frequency stimuli is supported by multiple channels, each tuned to a particular band (much like in a sound system's graphic equalizer). Evidence corroborating this idea comes from three kinds of psychophysical experiments. (1) Adaptation: exposure to a high contrast grating of a specific frequency reduces the sensitivity (that is, raises the detection threshold) for gratings of other spatial frequencies to an extent depending on the frequency difference. In the context of the variable frequency and contrast test grating shown here, adaptation would manifest itself as a notch in the boundary along which the grating fades into an apparently uniform field, situated at the frequency of the adapting stimulus: at the adapted frequency, more contrast is needed for the grating to be perceived. (2) Subthreshold summation: the extent to which two subthreshold (that is, imperceptible) gratings of differing spatial frequencies shown together combine to elicit a suprathreshold percept depends on the difference between their frequencies. (3) Masking: the threshold for a faint test grating is elevated by a high-contrast mask grating superimposed on it to the extent that their frequencies match. In all three cases, the effects fade when the difference in spatial frequency is larger than about one octave (a factor of two). This would not be the case, were the contrast information not processed by a set of independent mechanisms, each tuned to an octave-wide band of spatial frequencies.

common to a wide range of cognitive tasks. Some examples of such general-purpose computational building blocks of biological information processing are outlined below (see Appendix A for a brief overview of the relevant mathematical concepts).

2.2.1 *Function approximation*

In numerical analysis, function approximation is the problem of recovering the form of an unknown function from a set of given argument-value pairs. It has been noted that this generic problem description fits well the standard scenario of supervised learning (Poggio 1990). In vision, for instance, an observer may be exposed to various views of a given shape, then required to determine whether a test view belongs to the same object. This can be done by approximating an indicator function that encodes the appearance of the object in question in the space of all views of all possible objects, then evaluating it at the test view (Poggio & Edelman 1990). In motor control, the function to be learned may map the intended action to a vector of muscle activations, and so on for other cognitive tasks (Poggio 1990).

2.2.2 *Density estimation*

An idea that is computationally related to function approximation is the estimation of the probability density of some quantity of interest over the relevant variables. Such estimation may proceed in an unsupervised fashion, or combined with class information; it leads to powerful methods for statistical inference and decision making, via the mathematical apparatus of Bayes theory and related approaches. Statistical inference is widely acknowledged to be an indispensable tool for cognition, in areas ranging from vision (Kersten & Schrater 2000) to conceptual learning (Tenenbaum 1999) and language acquisition (Clark 2001), where the relevant domain may be the set of constructions acquired from a corpus, and the output of the inference procedure – a breakdown of the probabilities of each construction depending on the context.

2.2.3 *Dimensionality reduction*

Learning theorists know that for function approximation (in particular, density estimation) to work well it must be conducted over a low-dimensional domain: because the required number of data points grows exponentially with dimensionality, function approximation in high-dimensional spaces is intractable (Bellman 1961). The problem of dimensionality reduction, also known in cognition as feature detection, is increasingly often seen as fundamental in language (Landauer & Dumais 1997) and in vision (Intrator & Edelman 1997).

In some cases, these abstract problems and the principles and algorithms used to address them have been mapped onto the function of the brain and its circuitry, resulting in explanatory models that span all three levels of Marr's program. For example, in olfaction the anatomy and the physiology of the pathway leading from the sensory epithelium to the glomeruli in the olfactory bulb (Lancet 1991; Shepherd 1992) can be seen as filtering data through a bank of radial basis functions (Poggio 1990). This operation implements what is known to be a universal approximation algorithm (Hartman et al. 1990) that can be used in learning from examples (Poggio 1990). The same algorithmic approach can support visual object recognition, as demonstrated by the Chorus of Prototypes model (Edelman 1999), in which the stimulus is represented by its similarities to (processed) memory traces of past stimuli. Recent single-cell studies in the monkey found neurons that are broadly and redundantly tuned to particular object categories (Freedman et al. 2001) and that embody an ensemble representation of interobject similarities that is veridical with respect to the distal stimuli (Op de Beeck et al. 2001). Both these findings had been predicted by the Chorus of Prototypes model (Edelman 1998; Edelman 1999).

3. Dealing with structure: A special challenge?

To be relevant to language, the computational principles behind these findings must be extended to situations that require highly structured representations. Recent work in various areas of cognition has been pursuing such an extension. For example, in complex analogy tasks a similarity-based model performs very well when the distributed representations it uses are made to reflect the structure of the input (Plate 1995; Eliasmith & Thagard 2001). Likewise, in vision the Chorus of Fragments model (derived from the Chorus of Prototypes), which aims at dealing with structured objects and scenes (Edelman & Intrator 2003), is based on the twin principles of distributed representation by similarities (mentioned above) and of the use of visual space to anchor the various shape fragments (Edelman 2002), introduced next.

3.1 The role of space in representing structure

The idea that space should serve as a natural scaffolding for supporting structured representations, whose roots go back to the ancient mnemonic Method of Loci (Neisser 1976: 137), is stated forcefully in Wittgenstein's *Tractatus* (Wittgenstein 1961, proposition 3.1431):

> The essential nature of the propositional sign becomes very clean when we imagine it made up of spatial objects (such as tables, chairs, books) instead of written signs. The mutual spatial position of these things then expresses the sense of the proposition.

In vision, sorting shape cues by their location in the visual field goes a long way toward solving the binding problem in the representation of object and scene structure (Edelman 1999; Clark 2000; Edelman 2002; Edelman & Intrator 2003). In particular, various components of a scene or an object need not be bound to each other in any special manner, as long as each of them is bound to its proper location in the visual space, merely by virtue of its appearance there.

In neurobiology, the spatial scaffolding approach to the representation of visual structure is consistent with the omnipresence in the monkey inferotemporal and prefrontal cortex of *what+where* neurons, which are both shape-tuned (signaling *what* is the stimulus), and location-selective (signaling *where* it appears) (Rao et al. 1997; Op de Beeck & Vogels 2000). On a larger scale, the neural substrate of the perceptually defined external space may be the cortical surface itself, as indicated by the ubiquity of maplike representations (Gallistel 1990) in vision (Ward et al. 2002), olfaction (Joerges et al. 1997), and audition (Shamma 2001).

3.2 Spatial representations for language

Functional (problem-level) analogies between language and vision suggest various parallels between the manner in which structure is dealt with in these cognitive domains (Minsky 1985). For instance, the treatment of a sentence with an embedded relative clause may be compared to the processing of a scene with occlusion (Figure 2, left; additional parallels are illustrated and discussed in Figure 2, right). In recent years space has been conjectured to play a central role in neurolinguistics. Consider, for example, the notion of iconicity in syntax (Simone 1995): "... not only motor but also cognitive operations such as language, which do not appear to have any intrinsic spatial organization, are maintained in registration with spatial systems, and [...] this attention-requiring linkage confers a processing advantage" (Coslett 1999). The iconicity hypothesis is intimately connected to Construction Grammar (Fillmore 1985; Goldberg 1995): linguistic freezes or prefabs (Landsberg 1995) that are spatially (or, equivalently, temporally; cf. Figure 3) iconic become constructions when parameterized (Erman & Warren 2000). Psycholinguistic and neuropsychological evidence in support of linguistic iconicity has been recently reviewed in (Chatterjee 2001).

On the level of neurobiology, in those areas of the human brain that support language, the counterpart to the visual *what+where* neurons mentioned above may be *what+when* neurons, tuned to particular structures appearing in a particular sequence (as illustrated in Figure 3). The possible role of temporal response properties of neuron assemblies in implementing sequence-sensitive processing has been discussed by (Pulvermüller 2002); parallels between the brain representations of space and time in vision and in audition have been pointed out, among others, by (Shamma 2001).

4. Treatment of structure in computational cognitive linguistics

Examples of space-based representations, which abound both in vision and in language, show that the goal of structure-sensitive processing by a distributed architecture is less forbidding than commonly thought, and that it is already within reach of cognition-general principles and mechanisms. This section outlines a cognitive approach to language, which is based on these foundations, and which casts the relevant computational, algorithmic and implementational issues in cognition-general terms.

4.1 Computational approach: The Chorus of Phrases and Construction Grammar

When applied to language, the idea of a distributed representation of structure based on similarities to multiple structured exemplars, called the Chorus of Fragments in the setting of visual scene processing (Edelman & Intrator 2003), translates into a *Chorus of Phrases*: a redundant ensemble of potentially overlapping, mutually reinforcing phrase fragments that, as Langacker puts it, "motivate" the sentence they cover:

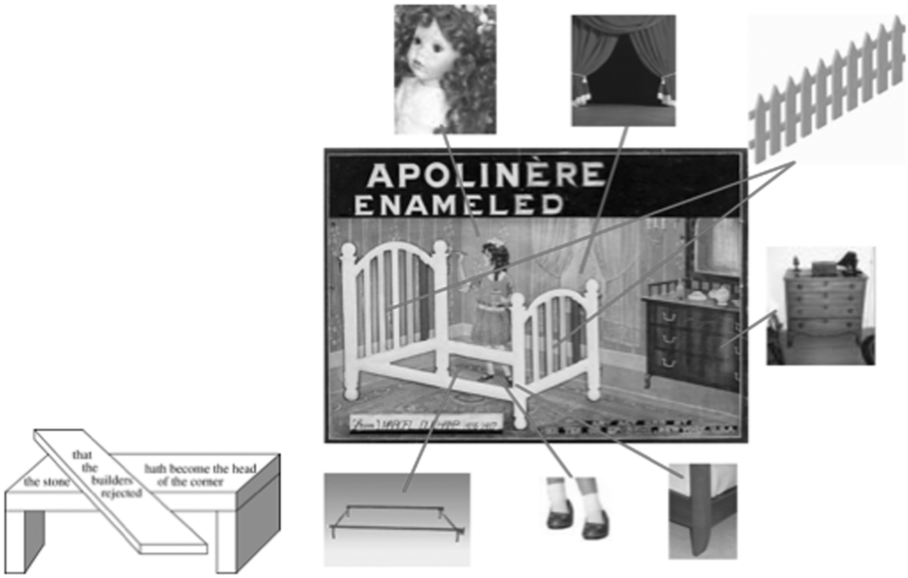


Figure 2. *Left:* there is a task-level analogy between interpreting a composite scene with occlusion and the processing of a sentence that contains an embedded relative clause (adapted from Minsky 1985: 269). The existence of such analogies between vision and language on the abstract level of the computational tasks (Marr & Poggio 1977) faced by the brain, along with the uniformity of the underlying low-level cortical mechanisms (Phillips & Singer 1997), suggests that cross-cognition commonalities should be sought also on the algorithmic level. *Right:* the postcard shown here (*Apolinère Enameled* by M. Duchamp) can be used to make the same point about parallels between language and vision (cf. the occlusion of the girl's legs by the bed frame), and more. For instance, Duchamp's painting could be represented (and understood) in terms of its local similarities to various familiar images (Chorus of Fragments (Edelman 2002; Edelman & Intrator 2003)), which need not match the scene perfectly (Edelman 1999); likewise, an utterance could be represented (and understood) in terms of the cloud of constructions (Chorus of Phrases (Solan et al. 2003b)) it evokes, as illustrated in Figure 3. Furthermore, just as many viewers fail to notice that the bed in this scene would be useless (look closely at the frame), subjects exposed to ungrammatical sentences of moderate complexity may rate them as no less felicitous than similarly structured grammatical ones. For example, the sentence "The apartment that the maid who the service had sent over was well decorated" tends to be rated as no worse (Gibson & Thomas 1999) – and, in some settings, better (Christiansen & MacDonald 2003) – than "The apartment that the maid who the service had sent over was cleaning every week was well decorated"; cf. (Gibson & Pearlmutter 1998; Chipere 1997; Chipere 2001).

> "... rather than seeing a composite structure as an edifice constructed out of smaller components, we can treat it as a coherent structure in its own right: component structures are not the building blocks out of which it is assembled, but function instead to *motivate* various aspects of it." (Langacker 1987: 453), italics in the original.

A schematic illustration of the Chorus of Phrases (CoPh) in action is shown in Figure 3, where a stimulus (which could be an entirely novel sentence) evokes a cloud of associations, pointing to snippets of previously encountered phrases, each of which approximately matches parts of the input, and which together cover all of it.

On the abstract, computational (Marr & Poggio 1977) level, the view of language as based on context-specific structural generalizations, exemplified by the CoPh approach, differs radically from that of generative theories such as the Minimalist Program (Lasnik 2002), which attempt to describe language in terms of universally valid syntactic rules projected by a categorically annotated lexicon. Recall that the basic theoretical challenge at the computational level is to specify what it is that needs to be done in the given task – in the present case, in language comprehension and production. According to the CoPh framework, comprehension involves constructing a distributed representation of the stimulus in terms of its structure-dependent similarities to multiple stored exemplars, which convey information both about form (the exemplars are, generally, patterns with slots; see Figure 3) and about meaning. Production consists of letting a set of exemplars chosen for their semantics interact and constrain each other until a fully specified linear sequence of terminals is ready for output.

This distributed approach, which does not distinguish between syntactic and semantic representations and processes, is broadly compatible with the tenets of the Cognitive school in linguistics (Langacker 1987), and, more specifically, with Construction Grammar (Fillmore 1985; Goldberg 1998; Goldberg 2003; Croft 2001). CoPh is, however, more than a mere metaphor for constructions, for several reasons. First, CoPh is deeply rooted in computational principles (multidimensional similarity spaces, distributed representations), algorithmic methods (statistical inference) and neural mechanisms (receptive fields and maps) that proved instrumental in analyzing other aspects of cognition. Second, by steering the goals of syntactic (and semantic) analysis toward those of cognition in general, CoPh brings to the fore a collection of mathematical tools hitherto not considered by linguists (see Appendix B: Mathematical tools). Third, an implemented model of language acquisition and processing situated within the CoPh framework provides empirical support and constraints for the construction grammar theories, as described briefly below.

4.2 Algorithmic and implementational issues: The ADIOS model

The algorithm behind this working model of language acquisition and processing (ADIOS, or Automatic DIstillation Of Structure) learns, in an unsupervised fashion, a streamlined representation of linguistic structures from untagged, large-scale natural language corpora (Solan et al. 2003b; Solan et al. 2004; Solan et al. 2005). The algorithm represents sentences as paths on a graph whose vertices are, initially, words. Significant patterns are defined as sets of paths in which a common prefix and suffix form a con-

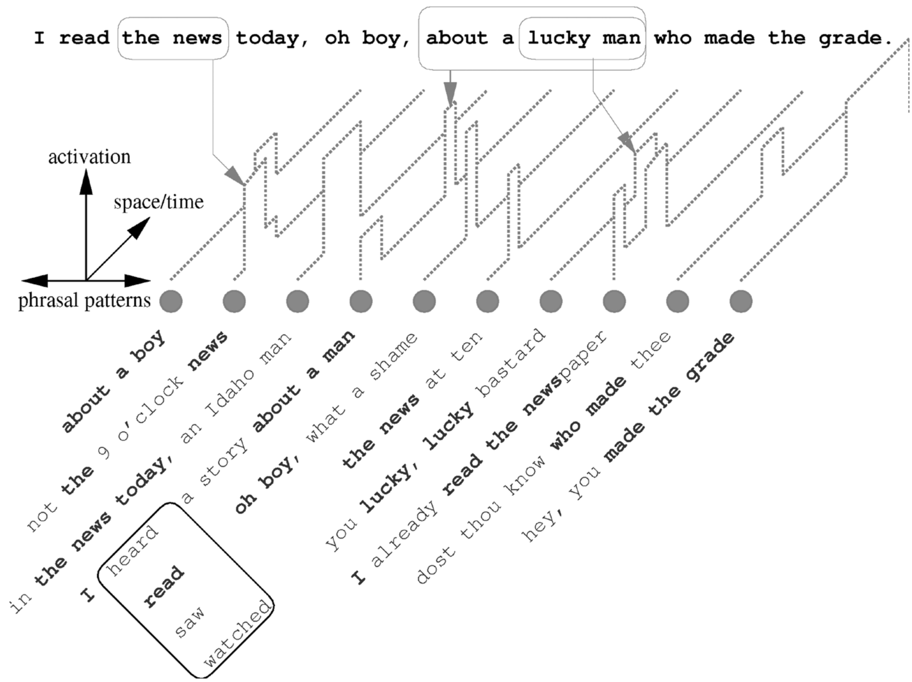


Figure 3. A schematic illustration of the Chorus of Phrases in sentence processing (for actual data from the ADIOS project, see Solan et al. 2003a; Solan et al. 2003b; Solan et al. 2005). An input sentence is shown along with a subset of phrases it evokes from memory, each of which matches some word, sequence of words, or, generally, a parameterized pattern (in cartouche: **I** [heard, **read**, saw, watched] a story **about a man**) in the input. The unfolding of each pattern's activation (which reflects its time-domain "receptive field") may be important (Pulvermüller 2002), but even without it the ensemble of active patterns is a highly informative representation, just as its counterparts in vision are (Edelman 1999; Edelman & Intrator 2003). The members of the ensemble disambiguate each other by supplying multiple interacting constraints on the interpretation. Consequently, it should be possible to process various queries about the input, both syntactic (voice, aspect, etc.) and semantic (thematic, connotational, conceptual). Moreover, it may be possible to use for that purpose generic cortical mechanisms (Phillips & Singer 1997; Maass et al. 2003) that would map the distributed phrase activation patterns onto the corresponding required outputs, as in the scenario of function approximation found across cognition (Poggio 1990; Intrator & Edelman 1997). From the computational standpoint, it is interesting to observe that one can reconstruct the input sentence itself, should that be required for some reason, from a number of sub-sequence (phrase or pattern) queries that is on the order of $n \log \alpha + \alpha \log n$, where n is the length of the sentence and α is the size of the lexicon (Skiena & Sundaram 1995). This computational complexity, which is quite benign in view of the α-fold parallelism inherent in a distributed lexicon, can be further reduced by allowing matching that is approximate (Jiang & Li 1996) in the sense of (Valiant 1984).

text surrounding a slot where locally distributionally equivalent (Harris 1954) elements may appear. In each iteration, such patterns, determined by context-sensitive statistical inference, form new vertices, and the graph is rewired, leading to the emergence of progressively more complex, hierarchically structured representations. The algorithm stops when no new patterns are found in a given iteration. Linguistic constructions thus correspond to trees composed of patterns and their associated equivalence classes. An entire utterance is typically represented by several such constructions (a Chorus of Phrases; cf. Figure 3), which may be activated to different degrees, depending on their fit to the input. Previously unseen inputs are processed by pursuing structural and lexical similarities to familiar patterns.

The probabilistic principle that drives the context-sensitive, hierarchical pattern abstraction process in the ADIOS model is related both to the notion of "suspicious coincidences" long thought to be the key to unsupervised learning in neural systems (Barlow 1959; Barlow 1989) and to the Minimum Description Length (MDL) criterion for representational efficiency (Bienenstock et al. 1997). Intuitively, two elements – such as two members of a potential linguistic construction or two fragments of a visual object – belong together to the extent that the probability of their joint appearance is higher than the product of the probabilities of their individual appearances; coding such elements as a unit results in a more concise representation. It has been conjectured that these principles, which can support structured learning in vision (Barlow 1990; Edelman et al. 2002a; Edelman et al. 2002b) and in language (Bienenstock et al. 1997; Clark 2001; Solan et al. 2003b), may provide "common foundations for cortical computation" (Phillips & Singer 1997). The ADIOS algorithm is based on a criterion for pattern unity that is specifically adapted to the sequential nature of language and to the graph-like data structure used to represent it. It is interesting to note that the entrenchment of pattern or construction-like units (the degree to which subjects treat them as such) depends on the corpus frequency of the corresponding subunit sequences (Harris 1998), which supports the notion that the patterns postulated by the ADIOS algorithms are psychologically real.

The implemented ADIOS model has been subjected to extensive tests, some of which focused on the acquisition of artificial languages generated by context-free grammars (CFG), and others – on learning from real natural language corpora such as CHILDES (MacWhinney & Snow 1985), ATIS (Moore & Carroll 2001) and the Bible (Resnik et al. 1999); only a few of these tests can be mentioned here. The CFG experiments involved two ADIOS instances: a teacher and a student. In each of the multiple runs, the teacher was preloaded with a ready-made context free grammar (using the straightforward translation of CFG rules into ADIOS patterns), then used to generate a series of training corpora with up to 6400 sentences, each with up to seven levels of recursion. After training in each run *i*, a student-generated test corpus $C^{(i)}_{learned}$ of size 10000 was used in conjunction with a test corpus $C^{(i)}_{target}$ of the same size produced by the teacher, to calculate precision and recall. This was done by running the teacher as a parser on $C^{(i)}_{learned}$ (precision measured by the teacher's acceptance of the student-generated sentences) and the student – as a parser on ${}_{target}{}^{(i)}$ (recall measured by the student's acceptance of novel sentences not seen during training).[5] The results – nearly 100% precision and about 95% recall – indicate a substantial capacity for unsupervised induction of context-free grammars even from very small corpora (Solan et al. 2005). Promising performance has also been demonstrated for real-life language. For example, a model trained on the CHILDES data attained a level of performance considered to be "intermediate" for 9th grade students when subjected to a standard test of English as a Second Language (ESL) proficiency (Solan et al. 2003b).

4.3 Select open questions

The proposed framework places within reach of empirical research many exciting open issues in language and cognition, some of which are outlined next.

4.3.1 *Linking psycholinguistics to visual psychophysics*

Recent psycholinguistic evidence indicates that listeners and readers routinely settle for "good enough" representations of the linguistic material they face, rather than seeking an exhaustive parse or even just a fully disambiguated semantic interpretation (Bever et al. 1998; Ferreira et al. 2002; Sanford & Sturt 2002); cf. Figure 2, right. A unified computational approach to vision and to language should help relate these findings to the cluster of phenomena in visual psychophysics known as "change blindness" (Simons & Levin 1997), which indicate that subjects do not fully parse visual scenes either.

4.3.2 *From the "big picture" to the neural mechanisms*

Marr and Poggio's framework calls for equal attention to the computation and the neurobiology-level understanding of cognition. On an abstract, computational level, construction-based approaches – in particular the Chorus of Phrases – readily integrate themselves into the rest of cognition, offering along the way a useful insight into the relationship between the final representational product of language and that of vision. Goldberg's thesis – "constructions all the way down" (Goldberg 2003) – can be taken to imply that the Chorus of Phrases evoked by an utterance or a text is just about all there is to its interpretation. There is a clear parallel between this stance and the conjecture that in vision the Chorus of Fragments is an adequate, and in fact the only reasonable, bottom line (Edelman 2002). On the level of mechanism, however, the details have yet to be worked out.

5. Defining performance in terms of sentence acceptance amounts to testing for the so-called "weak" generativity rather than the "strong" generativity, under which the derivation/parse trees are compared instead of sentences (Roberts & Atwell 2003). There is a good reason for this choice: any "gold standard" that can be used to evaluate strong generativity, such as the Penn Treebank (Marcus et al. 1994), invariably reflects its designers' preconceptions about language, which are often controversial among linguists themselves: the derivation trees are always stipulated, never observed. A learner who exhibits perfect weak generativity – that is, who accepts and produces all and only those sentences respectively produced and accepted by the teacher – is, for all practical purposes, perfectly successful.

4.3.3 *Between the mechanism and the virtual machine*

A crucial question is whether or not it is possible to avoid altogether the need to manipulate constructions dynamically, rather than through a prewired network. This is important because the act of binding a variable to a value (or inserting a constituent into a construction) dynamically is deeply problematic in the context of a neural implementation (Edelman & Intrator 2003). The ability of humans to do algebra or to program computers attests to the existence of some mechanism in the brain that at least creates the semblance of dynamic binding. People, however, must be trained for years before they become good at this kind of symbol manipulation, and it would be prudent to make it a means of last resort in a theory of any cognitive phenomenon that is more mundane than programming in Lisp. The same consideration applies to a related issue, recursion: although it has been recently reaffirmed by some linguists as the epitome of human uniqueness (Hauser et al. 2002), humans are notoriously bad at deep recursion (Gibson & Thomas 1999). In comparison, shallow recursion, as well as the manipulation of complexity controlled constructions, can be handled by finite means such as the ADIOS representation (Solan et al. 2003b). These considerations suggest that dynamic binding and deep recursion may both be supported in the brain by a virtual machine that is difficult to build and expensive to maintain and operate, and that is at least once removed from the neural mechanisms that are so good at supporting everyday cognition; cf. (Dennett 1991: 209).

5. Conclusions

Computational cognitive science holds that a comprehensive theory of any information processing task must lead to its understanding on several levels of abstraction (Marr & Poggio 1977). Although distinct, these levels cannot be studied independently, lest theorizing loses touch with psychological and neurobiological reality, or, conversely, the neurobiology becomes too myopic (Edelman 1999). Accordingly, the framework for computational cognitive linguistics outlined here is informed by the top-down computational and algorithmic principles of context-dependent probabilistic learning and is based on a bottom-up implementational scheme that is ubiquitous in the brain: computing with structured connections, which carry dynamically unfolding neural activation patterns and which support low-dimensional, distributed, redundant, graded representations. The vision of language it offers should boost cognitively oriented theories such as Construction Grammar (Croft 2001; Goldberg 2003) and help connect them to rich repositories of computational knowledge (from learning theory, probability and information, and natural language processing) and empirical data (from psychophysics and neuroscience) about the brain.

Acknowledgments

I thank Zach Solan, Eytan Ruppin, David Horn, Roni Katzir, Rick Dale and Bo Pedersen for stimulating discussions, and Michael Spivey, Shweta Narayan and two anonymous reviewers for constructive comments.

References

Aho, A. V., Hopcroft, J. E., & Ullman, J. D. (1974). *The Design and Analysis of Computer Algorithms*. Addison Wesley, Reading, MA.

Barlow, H. B. (1959). Sensory mechanisms, the reduction of redundancy, and intelligence. In *The mechanisation of thought processes* (pp. 535–539). London: H.M.S.O.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1,* 295–311.

Barlow, H. B. (1990). Conditions for versatile learning, Helmholtz's unconscious inference, and the task of perception. *Vision Research*, *30,* 1561–1571.

Baum, E. B. (2004). *What is thought?* Cambridge, MA: MIT Press.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton, NJ: Princeton University Press.

Bernárdez, E. (1999). Some reflections on the origins of cognitive linguistics. *Journal of English Studies, 1,* 9–27.

Bever, T. G., Sanz, M., & Townsend, D. J. (1998). The Emperor's psycholinguistics. *J. of Psycholinguistic Research, 27,* 261–284.

Bienenstock, E., Geman, S., & Potter, D. (1997). Compositionality, MDL priors, and object recognition. In M. C. Mozer, M. I. Jordan, & T. Petsche (Eds.), *Neural Information Processing Systems*, Vol. 9. Cambridge, MA: MIT Press.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. (1987). Occam's razor. *Information Processing Letters, 24,* 377–380.

Bod, R. (1998). *Beyond grammar: An experience based theory of language*. Stanford: CSLI Publications.

Campbell, F. W. & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *J. Physiol.* (*Lond.*)*, 197*, 551–566.

Chatterjee, A. (2001). Language and space: Some interactions. *Trends in Cognitive Sciences*, *5,* 55–61.

Chen, S. & Donoho, D. L. (1994). Basis pursuit. In *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers*, Vol. 1 (pp. 41–44). Pacific Grove, CA: IEEE Comput. Soc. Press.

Chipere, N. (1997). Individual differences in syntactic skill. *Working Papers in English and Applied Linguistics, 4,* 1–32.

Chipere, N. (2001). Native speaker variations in syntactic competence: Implications for first language teaching. *Language Awareness*, *10,* 107–124.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton & Co.

Christiansen, M. H. & MacDonald, M. C. (2003). Processing of recursive sentence structure: Testing predictions from a connectionist model. In preparation.

Churchland, P. S. & Sejnowski, T. J. (1992). *The Computational Brain*. Cambridge, MA: MIT Press.

Clark, A. (2000). *A theory of sentience*. Oxford: Oxford University Press.

Clark, A. (2001). *Unsupervised Language Acquisition: Theory and Practice*. PhD thesis, COGS, University of Sussex.

Coslett, H. B. (1999). Spatial influences on motor and language function. *Neuropsychologia*, *37,* 695–706.

Croft, W. (2001). *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press,

Cummins, R. (1996). *Representations, Targets, and Attitudes*. Cambridge, MA: MIT Press.

Damasio, A. R. & Tranel, D. (1993). Nouns and verbs are retrieved with differently distributed neural systems. *Proceedings of the National Academy of Science*, *90,* 4957–4960.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little, Brown & Company.

Edelman, S. (1998). Representation is representation of similarity. *Behavioral and Brain Sciences*, *21*, 449–498.

Edelman, S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Edelman, S. (2001). Neural spaces: A general framework for the understanding of cognition? A commentary on Shepard. *Behavioral and Brain Sciences, 24*, 664–665.

Edelman, S. (2002). Constraining the neural representation of the visual world. *Trends in Cognitive Sciences*, *6,* 125–131.

Edelman, S. & Christiansen, M. H. (2003). How seriously should we take Minimalist syntax? A comment on Lasnik. *Trends in Cognitive Science*, *7*, 60–61.

Edelman, S., Hiles, B. P., Yang, H., & Intrator, N. (2002a). Probabilistic principles in unsupervised learning of visual structure: Human data and a model. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems, 14* (pp. 19–26). Cambridge, MA: MIT Press.

Edelman, S. & Intrator, N. (1997). Learning as extraction of low dimensional representations. In D. Medin, R. Goldstone, & P. Schyns (Eds.), *Mechanisms of Perceptual Learning* (pp. 353–380). Academic Press.

Edelman, S. & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science*, *27*, 73–109.

Edelman, S., Intrator, N., & Jacobson, J. S. (2002b). Unsupervised learning of visual structure. In H. H. Bülthoff, C. Wallraven, S. W. Lee, & T. Poggio (Eds.), *Proc. 2nd Intl. Workshop on Biologically Motivated Computer Vision*, volume 2525 of *Lecture Notes in Computer Science* (pp. 629–643). Springer.

Eliasmith, C. & Thagard, P. (2001). Integrating structure and meaning: A distributed model of analogical mapping. *Cognitive Science*, *25,* 245– 286.

Erman, B. & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, *20,* 29–62.

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good enough representations in language comprehension. *Current Directions in Psychological Science*, *11*, 11–15.

Fillmore, C. J. (1985). Syntactic intrusion and the notion of grammatical construction. *Berkeley Linguistic Society*, *11*, 73–86.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The psychology of language*. New York: McGraw Hill.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*, 312–316.

Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.

Gibson, E. & Pearlmutter, N. (1998). Constraints on sentence comprehension. *Trends in Cognitive Sciences*, *2*, 262–268.

Gibson, E. & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*, 225–248.

Gibson, J. J. (1957). Survival in a world of probable objects. *Contemporary Psychology, 2,* 33–35.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (1998). Patterns of experience in patterns of language. In M. Tomasello (Ed.), *The new psychology of language* (pp. 203–219). Mahwah, NJ: Erlbaum.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*, 219–224.

Harris, C. L. (1998). Psycholinguistic studies of entrenchment. In J. Koenig (Ed.), *Conceptual Structures, Language and Discourse*, Vol. 2. Berkeley, CA: CSLI.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*, 140–162.

Hartman, E. J., Keeler, J. D., & Kowalski, J. M. (1990). Layered neural networks with Gaussian hidden units as universal approximations. *Neural Computation, 2*, 210–215.

Hauser, M., Chomsky, N., & Fitch, T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hume, D. (1740). *A Treatise of Human Nature*.

Iliopoulos, C. S. & Smyth, W. F. (1998). Online algorithms for covering. In *Proceedings of the 9th Australasian Workshop on Combinatorial Algorithms*, volume 6, 97–106.

Intrator, N. & Edelman, S. (1997). Learning low dimensional representations of visual objects with extensive use of prior knowledge. *Network*, *8*, 259–281.

Jiang, T. & Li, M. (1996). DNA sequencing and string learning. *Math. Syst. Theory*, *26,* 387–405.

Joerges, J., Küttner, A., Galizia, C. G., & Menzel, R. (1997). Representations of odors and odor mixtures visualized in the honeybee brain. *Nature*, *387,* 285–288.

Joshi, A. & Schabes, Y. (1997). Tree Adjoining Grammars. In G. Rozenberg & A. Salomaa (Eds.), *Handbook of Formal Languages*, Vol. 3 (pp. 69–124). Berlin: Springer.

Jurafsky, D. & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New York: Prentice Hall.

Kersten, D. & Schrater, P. R. (2000). Pattern Inference Theory: A probabilistic approach to vision. In R. Mausfeld & D. Heyer (Eds.), *Perception and the Physical World*. Chichester: John Wiley & Sons.

Knill, D. & Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Lancet, D. (1991). The strong scent of success. *Nature*, *351*, 275–276. News and Views.

Landauer, T. K. & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Landsberg, M. E. (Ed.). (1995). *Syntactic iconicity and linguistic freezes*. Berlin: Mouton de Gruyter.

Langacker, R. W. (1987). *Foundations of cognitive grammar*, volume I: Theoretical prerequisites. Stanford, CA: Stanford University Press.

Lasnik, H. (2002). The Minimalist Program in syntax. *Trends in Cognitive Science, 6,* 432–437.

Lodhi, H., Shawe Taylor, J., Cristianini, N., & Watkins, C. J. (2001). Text classification using string kernels. In T. K. Leen, T. G. Dietterich, & V. Tresp (Eds.), *Advances in Neural Information Processing Systems 13*, (pp. 563–569). Cambridge, MA: MIT Press.

Maass, W., Natschläger, T., & Markram, H. (2003). Computational models for generic cortical microcircuits. In J. Feng (Ed.), *Computational Neuroscience: A Comprehensive Approach*. Boca Raton, FL: CRC Press.

MacWhinney, B. & Snow, C. (1985). The Child Language Exchange System. *Journal of Computational Linguistics*, *12,* 271–296.

Manning, C. D. & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.

Marr, D. & Poggio, T. (1977). From understanding computation to understanding neural circuitry. *Neurosciences Res. Prog. Bull.*, *15,* 470–488.

Miller, G. (1962). Some psychological studies of grammar. *American Psychologist*, *17,* 748–762.

Minsky, M. (1985). *The Society of Mind*. New York: Simon and Schuster.

Moore, B. & Carroll, J. (2001). Parser comparison – context free grammar (CFG) data. online at http://www.informatics.susx.ac.uk/research/nlp/carroll/cfgresources/.

Mumford, D. (1994). Neuronal architectures for pattern theoretic problems. In C. Koch & J. L. Davis (Eds.), *Large scale neuronal theories of the brain* (pp. 125–152). Cambridge, MA: MIT Press.

Mumford, D. (1997). The mathematical modeling of cortical functioning and thought. In *Proc. Norbert Wiener Centennial Conference*. Providence, RI: Amer. Math. Society.

Nakano, Y., Felser, C., & Clahsen, H. (2002). Antecedent priming at trace positions in Japanese long distance scrambling. *Journal of Psycholinguistic Research*, *31*, 531–570.

Neisser, U. (1976). *Cognition and reality*. San Francisco: Freeman.

Ochsner, K. N., Chiu, C. Y. P., & Schacter, D. L. (1994). Varieties of priming. *Current Opinion in Neurobiology, 4*, 189–194.

Op de Beeck, H. & Vogels, R. (2000). Spatial sensitivity of Macaque inferior temporal neurons. *J. Comparative Neurology, 426*, 505–518.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Inferotemporal neurons represent low dimensional configurations of parameterized shapes. *Nature Neuroscience, 4,* 1244–1252.

O'Regan, J. K. (1992). Solving the real mysteries of visual perception: The world as an outside memory. *Canadian J. of Psychology*, *46,* 461–488.

Phillips, W. A. & Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences*, *20,* 657–722.

Plate, T. A. (1995). Holographic Reduced Representations. *IEEE Transactions on Neural Networks*, *6*, 623–641.

Poggio, T. (1990). A theory of how the brain might work. *Cold Spring Harbor Symposia on Quantitative Biology*, *LV,* 899–910.

Poggio, T. & Edelman, S. (1990). A network that learns to recognize three dimensional objects. *Nature*, *343,* 263–266.

Postal, P. M. (2004). *Skeptical linguistic essays*. Oxford/New York: Oxford University Press.

Pouget, A., Zemel, R. S., & Dayan, P. (2000). Information processing with population codes. *Nature Review Neuroscience*, *1*, 125–132.

Pulvermüller, F. (2002). A brain perspective on language mechanisms: From discrete neuronal ensembles to serial order. *Progress in Neurobiology*, *67*, 85–111.

Pulvermüller, F., Preissl, H., Lutzenberger, W., & Birbaumer, N. (1996). Brain rhythms of language: Nouns versus verbs. *Eur. J. Neurosci.*, *8*, 937– 941.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, *276*, 821–824.

Reitman, W., Nado, R., & Wilcox, B. (1978). Machine perception: what makes it so hard for computers to see? In C. W Savage (Ed.), *Perception and cognition: issues in the foundations of psychology*, volume IX of *Minnesota studies in the philosophy of science* (pp. 65–87). Minneapolis, MN: University of Minnesota Press.

Resnik, P., Olsen, M. B., & Diab, M. (1999). The Bible as a parallel corpus: annotating the 'Book of 2000 Tongues'. *Computers and the Humanities*, *33,* 129–153.

Rissanen, J. (1987). Minimum description length principle. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of Statistic Sciences*, volume 5 (pp. 523–527). J. Wiley and Sons.

Roberts, A. & Atwell, E. (2003). The use of corpora for automatic evaluation of grammar inference systems. In D. Archer, P. Rayson, A. Wilson, & T. McEnery (Eds.), *Proc. of the Corpus Linguistics 2003 conference*. UCREL, University of Lancaster.

Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346,* 174–177.

Sanford, A. J. & Sturt, P. (2002). Depth of processing in language comprehension: Not noticing the evidence. *Trends in Cognitive Sciences*, *6*, 382–386.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, *5*, 340–348.

Shepard, R. N. (1975). Form, formation, and transformation of internal representations. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 87–122). Hillsdale, NJ: Erlbaum.

Shepherd, G. M. (1992). Modules for molecules. *Nature*, *358*, 457–458. News and Views.

Simone, R. (1995). Iconic aspects of syntax: a pragmatic approach. In R. Simone (Ed.), *Iconicity in language* (pp. 153–169). Amsterdam/ New York: John Benjamins.

Simons, D. J. & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science*, *1*, 261–267.

Skiena, S. & Sundaram, G. (1995). Reconstructing strings from substrings. *Journal of Computational Biology*, *2*, 333–353.

Solan, Z., Horn, D., Ruppin, E., & Edelman, S. (2004). Unsupervised context sensitive language acquisition from a large corpus. In L. Saul (Ed.), *Advances in Neural Information Processing*, volume 16. Cambridge, MA: MIT Press.

Solan, Z., Ruppin, E., Horn, D., & Edelman, S. (2003a). Automatic acquisition and efficient representation of syntactic structures. In S. Thrun (Ed.), *Advances in Neural Information Processing*, volume 15. Cambridge, MA: MIT Press.

Solan, Z., Ruppin, E., Horn, D., & Edelman, S. (2003b). Unsupervised efficient learning and representation of language structure. In R. Alterman & D. Kirsh (Eds.), *Proc. 25th Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Tenenbaum, J. B. (1999). Bayesian modeling of human concept learning. In S. A. Solla, T. K. Leen, & K. R. Miller (Eds.), *NIPS* (*Advances in Neural Information Processing Systems*), volume 12. Cambridge, MA: MIT Press.

Tomasello, M. (1998). Introduction: A cognitive-functional perspective on language structure. In M. Tomasello (Ed.), *The new psychology of language* (pp. vii–xxiii). Mahwah, NJ: Erlbaum.

Tulving, E. & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247*, 301–306.

Valiant, L. G. (1984). A theory of the learnable. *Communications of the ACM, 27,* 1134–1142.

Ward, R., Danziger, S., Owen, V., & Rafal, R. (2002). Deficits in spatial coding and feature binding following damage to spatiotopic maps in the human pulvinar. *Nature Neuroscience*, *5*, 99–101.

Wiggs, C. L. & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.*, *8,* 227–233.

Wilson, H. R. & Bergen, J. R. (1979). A four-mechanism model for threshold spatial vision. *Vision Research*, *19*, 19–32.

Wittgenstein, L. (1961). *Tractatus Logico-philosophicus* (trans. D. F. Pears & B. F. McGuinness). London: Routledge.

Zadrozny, W. (1994). From compositional to systematic semantics. *Linguistics and philosophy*, *17*, 329–342.

Appendix A: Some useful mathematical concepts

The following is a brief glossary of some of the mathematical concepts used in this chapter to describe the state of the art in the computational understanding of cognition.

Functions

Much of the essence of the idea of learning in cognition is captured by the mathematical notion of a function. Formally, a function is a mapping (a specification of correspondence) from one set, which is called the domain, to another, called the range. In the example of Figure 4, left, the function f maps the element $a1$ in the domain A to $b2$ in the range B, $a2$ to $b1$, and so on. The mapping that defines a function must be unequivocal in that a given element in the domain cannot be mapped to more than one element

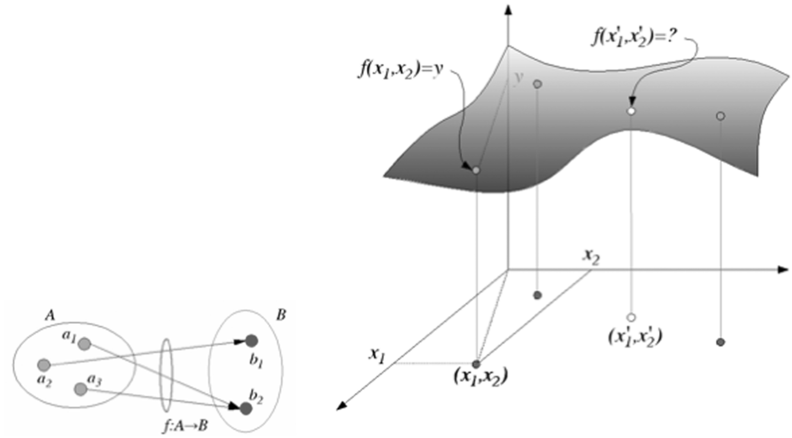


Figure 4. Illustrations of the basic idea of a function (*left*) and of learning a function from examples (*right*); see text for explanations.

in the range (although any number of elements in the domain can be mapped to the same one in the range, as illustrated in Figure 4, left). Thus, every time a cognitive system learns to attach a valence to a stimulus or associate it with a response, it learns a function defined over the set of possible stimuli. Note that the sets in question may include variables that are internal or external to the system, and are, in general, multidimensional spaces (see below).

Multidimensional spaces

A point in an n-dimensional space can be thought of as an ordered collection of n independent measurements (feature values) that define some entity of interest. For example, the state (configuration) of a human arm can be described by four numbers: three to specify the angular position of the spherical shoulder joint and one to specify the elbow angle (including the hand introduces many more degrees of freedom and hence many more dimensions). Thus, each configuration of the arm corresponds to a point in a four-dimensional space, and learning to position an arm amounts to learning a function from the space of the relevant muscles to this configuration space.

Learning and function approximation

In this conceptual framework, learning from examples amounts to interpolating an unknown function from the given correspondences between some of the elements in its domain and its range (Poggio 1990). In the illustration shown in Figure 4, right, the function f maps a two-dimensional space (the horizontal plane) into a one-dimensional space (the vertical axis); it can be thought of as a surface defined over the plane. Given the height of the surface (the value of the function) at several points (filled circles), one can try to determine its value at a new point (open circle). In the context of controlling an arm, for instance, this would amount to a generalization of the previously available control settings to determine a setting for a new target configuration. The mathematical, psychophysical, and neurobiological aspects of this approach as applied to visual object recognition are described in (Edelman 1999).

Dimensionality reduction

The mathematical tools associated with the concept of multidimensional spaces can be applied to the description of brain states and of their evolution over time (Mumford 1994; Mumford 1997; Edelman

2001). In the most straightforward fashion, one dimension is assigned to describe the activity level of each neuron (Churchland & Sejnowski 1992). This results in a space with many billions of dimensions; apart from the convenience of the mental picture of a brain state as a point in such a space, not much is gained, because of the overwhelming computational complexity of dealing with such high-dimensional spaces. Because certain kinds of brain representations are necessarily high-dimensional,[6] cognition must involve, at various stages, the reduction of dimensionality to a level that is computationally manageable. The dimensionality of a representation can be reduced by projecting it into a lower-dimensional space; for details and applications of this idea, see (Edelman & Intrator 1997). Note that such a projection is a function, which can be learned from examples (as described above).

Probabilities and statistical inference by density estimation

The most knowledge one can possess about any situation that involves information processing is the joint probability of all the relevant variables, $P(X1, X2, \ldots, Xn)$. This profound insight can be traced back to the writings of David Hume:

> "All kinds of reasoning consist in nothing but a comparison, and a discovery of those relations, either constant or inconstant, which two or more objects bear to each other."
> (Hume 1740, Part III, Sect. II)

> "An experiment loses of its force, when transferr'd to instances, which are not exactly resembling; tho' 'tis evident it may still retain as much as may be the foundation of probability, as long as there is any resemblance remaining." (Part II, Sect. XII)

> "... all knowledge resolves itself into probability..." (Part IV, Sect. I)

Hume's realization of the central and crucial role of *statistical inference* in knowledge generation (that is, learning) has been developed by many others, including his contemporary Thomas Bayes, the pioneering statisticians Karl Pearson and Ronald A. Fisher, and the neurobiologist Horace B. Barlow. Their combined insights led to the modern applications of inference to vision and other senses (Barlow 1990; Knill & Richards 1996), as well as to language (Manning & Schütze 1999).

The conception of visual learning as inference is naturally complemented by the emerging view of perception as statistical decision making, stated cogently in the following passage by the originator of the ecological theory of perception, the psychologist J. J. Gibson:

> "... the percept is always a wager. Thus uncertainty enters at *two* levels, not merely one: the configuration may or may not indicate an object, and the cue may or may not be utilized at its true indicative value." (Gibson 1957)

As a simple concrete example, consider a perceptual system that monitors the values of three features, $X1$, $X2$ and $X3$, which are related to the presence of four possible objects as coded by $X4 = \{O1, O2, O3, O4\}$ seen at one of two possible angles, $X5 = \{A1, A2\}$. In this case, the knowledge of the joint probability density $P(X1, X2, X3, X4, X5)$ would allow the system to estimate the conditional probability of each combination of object and angle, given the observed feature values: $P(X4, X5 | X1, X2, X3)$. This information, in turn, would suffice to support optimal decision making on the basis of the maximum likelihood criterion (which combination of values of $X4$ and $X5$ is most likely in the light of the measurements?). It is important to note that the same tools used to reason about – and to learn – a function from examples are also applicable to probability densities.

6. For example, the retinal signal, whose raw dimensionality runs in the millions (it is equal to the number of axons in the optic nerve), or the motor control signal, whose dimensionality is at least the same as the number of distinct muscles in the body.

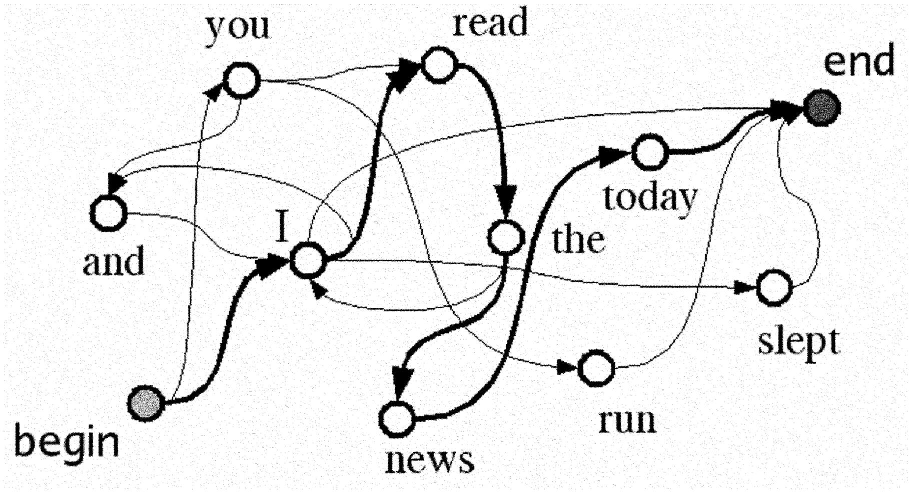


Figure 5. A very simple graph of the kind used as the basic data structure by the ADIOS algorithm

The graph data structure

The kind of data that arise in the context of natural language processing, namely, sequences defined over a finite alphabet or lexicon of discrete symbols, can be captured by graphs – a useful data structure that is extensively studied in computer science (Aho et al. 1974). Formally, a graph is specified by two sets: the vertices, and the edges (which may be directed) that connect some of them. Figure 5 illustrates a simple graph whose vertices are labeled by the symbols {begin, end, I, and, you, read, news, the, run, today, slept}. Traversing some of its directed edges while writing down the encountered vertices yields sentences such as "I read the news today", "you run", "I slept", and "you and I slept". The ADIOS algorithm outlined in Section 3.2 uses a graph of this type as its basic data structure.

Appendix B: Mathematical tools for computational cognitive linguistics

From the standpoint of methodology, a computationally motivated approach to cognitive linguistics does not imply pitching "mathematics versus psychology" – an expression used by Tomasello as a section heading in his introduction to *The New Psychology of Language* (Tomasello 1998: ix). Rather, the usual tools of mathematical linguistics (such as formal languages, λ-calculus and symbolic logic) should be supplemented by new ones. Some of these, which proved well suited for analyzing distributed representations in various areas of cognition, are listed below.

Syntax: From constituent trees to string cover

Computational learning theory offers various tools capable of dealing with distributed, potentially overcomplete (Chen & Donoho 1994) representations of sequence data. One of these is string kernels, a representation that tallies the occurrences of specific symbols in specific locations, and supports reasoning about global properties of the sequence probed in this manner and, in particular, about features that can help classify it (Lodhi et al. 2001). Similar methods are increasingly in demand in computational biology, because of the sequential nature of the data in both domains, and, specifically, because of the close analogy between text analysis by the identification of multiple, overlapping local patterns on the one hand, and hybridization approaches to DNA sequencing on the other hand. Recent developments in this field include derivations of the algorithmic complexity of specifying a string by its substrings (Skiena & Sundaram 1995; Jiang & Li 1996; Iliopoulos & Smyth 1998). This approach is distinct from (and more relevant for our present purposes than) treebank based parsing (combining multiple local or partial parse trees (Joshi & Schabes 1997; Bod 1998)) in that the cover it seeks need be neither precise nor exclusive.

Semantics: From functions to constructions and relations

According to the Chorus of Phrases metaphor (Figure 3), the representation of a sentence by an ensemble of active units can be approximately described as a relation (namely, as the subset of units whose activity exceeds some threshold). As in Construction Grammar (Goldberg 2003), this representation captures both semantic and syntactic information about the input. Interestingly, recent work in computational semantics addressing various problematic aspects of compositionality suggests that systematicity of meaning is better served by defining meaning as a relation over sentence parts (Zadrozny 1994), rather than as a function of the parts, as stipulated by the classical, Fregean approach.

Acquisition: From parameter setting to structure discovery

The ascendancy of the generative grammar and its accompanying innateness postulate over competing distributional and behaviorist ideas in the 1960s can be ascribed in a large part to the inadequacy of the contemporary statistical inference methods and the perceived inability of association-based learning to handle recursion. Statistics, however, need not be limited to counting word frequencies: in computer science, the integration of advanced statistical inference (including Bayesian methods, the Minimum Description Length principle and other related information-theoretic tools), progress in computational learning theory, efficient algorithms, and cheap hardware led to important conceptual progress, as well as to practical achievements (Manning & Schütze 1999). Likewise, learning need not be limited to the establishment of pairwise associations: bounded-depth recursively structured patterns can be learned from examples, by efficient algorithms that rely on modern statistical inference (Solan et al. 2003b; Solan et al. 2004); see (Clark 2001) for an overview of the recent progress in this field.