Sergio Scalise
Elisabetta Magni
Antonietta Bisetto
Editors

Studies in Natural Language and Linguistic Theory 76

Universals of Language Today

Springer

Scalise · Magni · Bisetto *Eds.*
Universals of Language Today

UNIVERSALS OF LANGUAGE TODAY

Studies in Natural Language and Linguistic Theory

VOLUME 76

For other titles published in this series, go to
www.springer.com/series/6559

UNIVERSALS OF LANGUAGE TODAY

Edited by

SERGIO SCALISE

University of Bologna, Italy

ELISABETTA MAGNI

University of Bologna, Italy

and

ANTONIETTA BISETTO

University of Bologna, Italy

Springer

Editors
Sergio Scalise
via Cartoleria, 5
40124 Bologna (BO)
Università di Bologna
Italy
scalise@lingue.unibo.it

Elisabetta Magni
via Cartoleria, 5
40124 Bologna (BO)
Università di Bologna
Italy
elisabetta.magni@unibo.it

Antonietta Bisetto
via Cartoleria, 5
40124 Bologna (BO)
Università di Bologna
Italy
antonietta.bisetto@unibo.it

Front cover: Castanea sativa, Photograph supplied by J.R.P. van Hoey Smith, Arboretum Trompenburg, Rotterdam, The Netherlands

ISBN: 978-1-4020-8824-7 e-ISBN: 978-1-4020-8825-4

Library of Congress Control Number: 2008932625

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Edoardo Vineis
amicitiae tuae memores atque grati

Preface

This book collects the contributions presented at the international congress held at the University of Bologna in January 2007, where leading scholars of different persuasions and interests offered an up-to-date overview of the current status of the research on linguistic universals.

The papers that make up the volume deal with both theoretical and empirical issues, and range over various domains, covering not only morphology and syntax, which were the major focus of Greenberg's seminal work, but also phonology and semantics, as well as diachrony and second language acquisition.

Diverse perspectives illustrate and discuss a huge number of phenomena from a wide variety of languages, not only exploring the way research on universals intersects with different subareas of linguistics, but also contributing to the ongoing debate between functional and formal approaches to explaining the universals of language.

This stimulating reading for scientists, researchers and postgraduate students in linguistics shows how different, but not irreconcilable, modes of explanation can complement each other, both offering fresh insights into the investigation of unity and diversity in languages, and pointing to exciting areas for future research.

- A fresh and up-to-date survey of the present state of research on Universals of Language in an international context, with original contributions from leading specialists in the field.
- First-hand accounts of substantive findings and theoretical observations in different subareas of linguistics.
- Huge number of linguistic phenomena and data from diffferent languages analyzed and discussed in detail.
- Essential reading for scientists, researchers and postgraduate students in linguistics.

Università di Bologna
Italy

Sergio Scalise
Elisabetta Magni and Antonietta Bisetto

Contents

Contributors

Geert Booij
University of Leiden

Guglielmo Cinque
Ca' Foscari University, Venice

Greville G. Corbett
Surrey Morphology Group

William Croft
University of New Mexico

Denis Delfitto
University of Verona

Matthew S. Dryer
University at Buffalo

Emiliano Guevara
University of Bologna

Martin Haspelmath
Max-Planck-Institut für evolutionäre Anthropologie, Leipzig

Aditi Lahiri
University of Oxford

Elisabetta Magni
University of Bologna

Frans Plank
Universität Konstanz

Anna Giacalone Ramat
University of Pavia

Paolo Ramat
University of Pavia

Sergio Scalise
University of Bologna

Irene Vogel
University of Delaware

Introduction

Since Greenberg's 1963 seminal work, research has constantly provided fresh insights into the exploration of unity and diversity in the realm of languages.

As is well known, a major idea that emerges from the 45 universals is that cross-linguistic variation is to some extent limited by the fact that certain items and structures in language systems are dependent on one another. And in its quest for the bounds of linguistic variability, the study of linguistic universals is closely linked to typological research, which defines and compares the traits and the patterns that languages share, producing the data upon which cross-linguistic generalizations are made.

Four decades of systematic language sampling and investigations greatly contributed to ascertain dependency relationships and co-variation between elements across languages, whereas considerable effort has also been expended to seek explanations for the way universals, both absolute and implicational, intervene to shape the limits of linguistic diversity. As a result, an enormous amount of empirical data and literature provides present-day linguistics with multiple perspectives and issues concerning those phenomena and correlations that in world's languages significantly occur "with more than chance frequency".

This very expression inspired the international congress held at the University of Bologna in January 2007, where contributions presented by leading scholars offered an up-to-date overview of the current status of the research on linguistic universals as instantiated in works that deal with both theoretical and empirical issues, and illustrate diverse modes of explanation.

Greenberg's insights have been incorporated in much linguistic work, not only in functional-typological approach but in generative-formal approach as well. And programmatically, the conference brought together linguists of different persuasions and interests, in order to determine to what extent the various kinds of explanation complement each other, and to what extent they can shed light on the interplay between research on universals and recent developments in the different areas of lin-

---

This volume is dedicated to the memory of our friend and colleague Edoardo Vineis, who was one of the organizers of the Conference on the Univerals of Language but passed away suddenly in August 2007.

guistics.[1] In fact, the articles collected in this volume range over different domains, covering not only morphology and syntax, which were the major focus of Greenberg's inspiring article, but also phonology and semantics, as well as diachrony and second language acquisition.

The overall trends of studies undertaken by scholars working with contemporary concepts of language typology reveal that the methods of analysis have greatly evolved, and the goals have correspondingly changed too.

Although Greenberg was quite cautious about the relevance of type classifications, and the import of correlations involving word order patterns, in the second half of the last century a great amount of literature consolidated the tradition concerning holistic language types, and a number of scholars managed to formulate hypotheses on implications, both speculating about idealized holistic principles (e.g. Lehmann 1973, 1978, Hawkins 1983), and exploring the interconnections between word-order and other structures (among others, Stassen 1985, Siewierska 1996, Haspelmath 1997).

The idea of 'deeper' explanations for word order universals turned out to be very attractive for generative linguists as well, who endeavored to discover broad correlations with underlying macroparameters, and overall motivations rooted in acquisition processes. Also, within the generative grammar framework many attempts have been made to refer some of the 45 universals to the notion of 'parameter', more specifically to the 'head-complement' parameter (Graffi 1980, Stowell 1981, Koopman 1983, Travis 1984, Coopmans 1985, Giorgi and Longobardi 1991).

In recent years, however, external critics and also insiders in both theoretical frameworks repeatedly observed the discrepancy between such ambitious purposes and the actual results. Evidently, these earlier objectives concerning broad cross-domain correlations and large-scale connections were not reachable: "It is illusory to think that linguistic diversity can be captured by a few holistic types, or a few word-order types, or a few parametric switches" (Haspelmath 2006: 17).

In consonance with this, in current research on universals the targets and the desired outcomes seem to be somewhat more restricted, since those generativists interested in grammar comparison tend to focus on microparameters (e.g. Kayne 1994) rather than on macroparameters, whereas functionalists tend to rely on correlations holding between elements in closely related constructions, rather than on large-scale implications (Haspelmath 2006).

And a notable shift of interest is observable in typology as well, which over the past decade has matured into a full-fledged discipline, with different objectives and with its own theories: "the new goal of typology is the development of theories that explain why linguistic diversity is the way it is" (Bickel 2007: 239).

The increasing appreciation of linguistic diversity in its own right amounts to a growing interest in geographical and historical investigations in order to explain

[1] In a similar vein, some recent works offer a balanced viewpoint in comparing and contrasting functional and formal perspectives in different areas: among others, Darnell et al. (1999) and Fischer (2007).

both the areal patterning of typological features and the qualitative and quantitative interrelations between typological distributions (e.g. Nichols 1992, Haspelmath et al. 2008). Accordingly, the essentials of contemporary typology are: "framework-neutral definitions, emphasis on codability in definitions and in applications of theory, bottom-up or data-driven constructs, and concern with observable phenomena that pattern interestingly in the world's languages (in their frequency, their interaction with other parts of grammar, their geography, their history, etc.)" (Nichols 2007: 231).

Notwithstanding the attraction of the fascinating panorama of linguistic variation that unfolds in space and time, typological research does not disregard the general principles that underlie the superficial variety and shape its complexity. In fact, a number of overarching and particular issues concerning their finding and explanation are currently under debate, thus confirming that research on universals of language is still alive and well.

The papers collected here also focus on some major problems and perspectives, which will be briefly discussed in the following paragraphs.

1 Universals and Grammatical Categories

Typological investigations have been largely based on the implicit assumption of a number of universal grammatical categories, relations and constructions, which are derived from the Latin and English grammatical tradition, notions such as parts of speech, passive, subject, direct object, relative clause, etc. The increasing attention devoted to the structural diversity displayed by human languages, however, entails a constant reassessment of existing taxonomies. And also, linguistic variability again and again demonstrates that these basic notions are extremely difficult to define in both formal and cross-linguistically valid terms (see the extensive discussion in Baker 2003).

As a consequence, the view that structural categories of grammar are universal has been variously interpreted, and more recently disputed by many linguists. The relevance of categorial concepts for the practice of linguistic typology is specifically focused in Dryer (1997), where the author posits the question whether grammatical relations are to be held either as fundamentally cross-linguistic or fundamentally language-particular concepts.

His claim that, from a functionalist view of language, the universality of structural categories is "at most a convenient fiction" (Dryer 1997: 117) is endorsed in Croft (2000) and (2001), where language-specific categories are assumed to reflect the prevalence of different and competing functional principles within the same grammatical domain.[2]

Strong arguments for radical language-particularity are also provided in Haspelmath's (2007) article, which explicitly argues against the existence of pre-established

[2] See also the discussion in Cristofaro (forth).

categories. The awareness that categories are language-particular, he says, has important consequences for both language description and typology: the latter in particular should abandon category-based comparison in favor of substance-based comparison.

Conversely, in the same special issue of *Linguistic Typology*, Newmeyers' (2007) response to this proposal points out that, even granting the assumption that semantic concepts are universal, semantically based categorial assignments in language description are just as problematic as form-based assignments. Moreover, he maintains that cross-linguistic formal categories are empirically motivated and, indeed, necessary to typology.

With respect to parts of speech, also the detailed survey in Schachter and Shopen (2007) assumes that the primary criteria for their classification are not semantic, but grammatical: namely, properties like the word's distribution, its range of syntactic functions, and the morphological or syntactic categories for which it is specifiable.

In this volume, a further contribution to this crucial debate is provided by **Ramat**, whose balanced discussion starts from the fundamental role of categorization with respect to cognition processes, and looks for convergences in the somewhat sterile divide between universal formalism vs. relativistic functionalism.

Since the relevant properties of a category belong to different levels of the linguistic organization (semantic-functional, morphosyntactic, pragmatic, etc.), the author suggests that different criteria, not only formal, may help to make correct categorial assignments.

Furthermore, in spite of those obvious considerations, which seem to lead to an absolute relativity of categorial concepts, Ramat maintains that parts of speech can be considered as universal categories and good tools of analysis for all languages. Provided that the distinction is made between categories as theoretical concepts, abstractions used in linguistic analysis, and their practical, real implementations in natural languages.

2 Universals and Language Variability

In recent years, the emphasis placed on linguistic diversity promoted the constant refinement of the tools and the methods for language description, which entailed also the growing awareness that grand type notions are inadequate to capture the kaleidoscopic variability of human languages.

Accordingly, among modern typologists individual structural patterns, or large sets of fine-grained variables are preferably adopted as *comparanda*, whereas the traditional four-way morphological classification of languages, is generally ignored, if not overtly criticized.

In a similar vein, **Haspelmath's** paper specifically discusses the validity of the agglutination/fusion distinction by making explicit a number of implicit assumptions of earlier work, and by exposing these postulates to rigorous empirical testing. The use of these terms implies in fact assumptions concerning both the putative

homogeneity of the type within all areas of morphology, and the consistency with respect to three possibly correlating parameters: the Cumulation index, the Alternation index and the Suppletion index.

Yet, according to the scrutiny of the nominal and verbal inflectional morphology in a sample of 30 languages, the key-features of the composite types 'agglutinating' and 'fusional' do not show statistically significant correlations, and the results are thus largely negative with respect to the validity of the agglutination-fusion distinction.

Considering that, on the other hand, these parameters do show relevant values and effectively co-occur in Turkish and Latin, the so-called Agglutination Hypothesis seems just an expedient to capture the eccentricity of certain non-Indo-European languages, whereas Latinocentrism traditionally informed the 'ideal' type of fusional language.

Haspelmath's study thus presents a strong challenge to those scholars who still want to maintain the traditional terminology. What's more, it not only shows that future work on morphological typology will need to be more careful, but also gives empirical evidence that holistic typology and gross morphological types, insofar as they apply to entire languages, fail the test when confronted with the intrinsic variability that exists within and across human languages.

In the same perspective, the real scope of universals can be questioned as well. **Lahiri and Plank** address explicitly this crucial issue in a rich contribution that centers on the relationship between universals and language variability, as it results from individual lexical and grammatical innovations. Extensive evidence from the syntax, morphology, and phonology of many languages proves that coherent sets of entities do frequently violate a universal that remains true of the respective languages as a whole. As a consequence, universals are not perforce and axiomatically to be predicated of entire languages, and it is thus sensible to distinguish between different types of universals, depending on which kinds of representations they specifically constrain (Plank 2007).

On the other hand, if true universals of language are only those principles that pervasively and consistently influence all parts and all representations of forms and constructions, then we are forcibly left with a very few basic constraints that limits individual variations and linguistic diversity.

3 Universals and Phonology

The above-mentioned paper devotes plenty of space also to the reflection on universals and phonology. Admittedly, a fruitful typological approach to phonology cannot be confined to questions of phoneme inventories, but must also include cross-linguistic comparison on the basis of the relationship between the linguistic system and its external (phonetic) manifestation (Croft 2003: 19). In this regard, the further advantages of looking at feature systems are made clear in Lahiri's excellent discussion concerning asymmetry in phonological systems.

The data provide evidence that asymmetry is inherent to features, feature distributions, and the direction of phonological rules; furthermore, the pervasiveness of featural asymmetry also extends to speech perception. Her analysis of this universal tendency is a fascinating illustration that avoids the possible unrepresentativeness of cross-linguistic data that language universals investigators usually work with, and also corroborates the view that the explanation for universals in phonology is likely to be found in articulatory, acoustic and also auditory phonetics.[3]

While phonology, as others areas of linguistics, is able to look at universals and typology with four decades of perspective, with respect to the study of prosodic structures systematic cross-linguistic investigation really started only about 20 years ago, when a full model of prosodic phonology was developed.

The core aspects of this sub-discipline are carefully outlined in **Vogel's** paper, which proposes ten possible universals of prosodic phonology by addressing both issues of abstract representation as well as the appearance of surface patterns. The article thus offers a well-articulated and detailed discussion of the interaction of formal and typological approaches to the study of the prosodic structure of language. The author considers the import of counter-examples, paying due attention both to the possibility of turning putative absolute universals into tendencies, and to the alternative of explaining them by more abstract (but independently justifiable) analyses. This pathbreaking work paves the way for further research and testing, which will certainly help to discover if additional properties of prosodic structures can qualify as universals of prosodic structure.

4 Universals and Morphology

In spite of its pioneering role in nineteenth century typology, in the second half of the last century morphology received relatively little attention, and still, the traditional taxonomies focused on the degree and the kind of morphology found across languages, are the idea that probably comes to mind first on hearing the term 'morphological typology'. Other aspects of the sub-discipline, on the other hand, seem to play indeed a minor role in typological studies, where morphology, as interface component of grammar par excellence, more often serves as the context for phonological processes, or as a diagnostic for the presence of morphosyntactic features (Baerman and Corbett 2007: 115).

This is probably due to the fact that morphology is in many respects a language-specific area, and so is even the very presence of a meaningful morphological component. Hence, in this field typologists are confronted with genuine difficulties in finding generalizable parameters of variation and organizing principles that are valid cross-linguistically.

Some recent typological surveys of word formation processes (Aikhenvald 2007) and inflectional morphology (Bickel and Nichols 2007), however, suggest that mor-

[3] On this point see also the discussion in Croft (2003: 117–120).

phological typology is presently benefiting from conspicuous theoretical advances in the sub-discipline. A deeper comprehension of the crucial notion 'word', new insights into the nature of allomorphy and paradigms, as well as the increasing interest in the diachronic, functional and cognitive dimensions of morphology brought by grammaticalization studies,[4] have now made it possible to provide typological investigations with firmer foundations and tools.

Three articles in this volume fully substantiate this renovated interest in morphological typology that, apart from holistic types, also implies a variety of dimensions along which we can typologize the world's languages[5].

By discussing the non-interruptability and non-accessibility of word-internal structure implied by a formal universal, the Lexical Integrity Principle, **Booij's** paper offers an important contribution to our understanding of the notion 'word' in grammar. This article investigates *to what extent* the word is an atomic unit with respect to other grammatical levels, in particular to syntax and semantics.

On the basis of telling examples, the author argues that, while non-interruptability can be maintained as a defining property of canonical wordhood, non-accessibility, on the other hand, should be rejected on empirical grounds. Cases of real violation of non-interruptability are cross-linguistically marginal phenomena, but there is no doubt that the different modules of the grammar cannot be excluded from having access to word-internal structure. Therefore, the non-interruptability part of the Lexical Integrity Principle must be conceived as the default situation for natural languages, rather than as an absolute universal that forbids syntactic manipulation.

Absolute morphological universals are difficult to find probably because the language-specificity of morphological processes reflects the language-specificity of the lexicon, from which grammaticalization derives the materials for word formation.

With respect to compound formation, **Guevara and Scalise** approach the search for universals using a lexical database that is much more sophisticated in its classificatory parameters than most previous work on compounding.

This article attempts to advance both empirically and analytically the study of compounding, the latter for instance in its honest discussion of some putative universals. The idea of a head-parameter, for instance, offers a fruitful typological perspective, but the empirical data do not fully support this hypothesis. In the same vein, other alledged universals (e.g. the right hand head rule, the root-compounding parameter, etc.), are seriously questioned.

Compounding phenomena are thus another area where the behavior of the world's languages seems to be governed by tendencies, rather than by universals of an absolute nature. But the domain of word formation, on the other hand,

---

[4] See, for instance, Anderson (2005: 55): "[s]ome things that we find in the morphology of a language are there not because the language faculty requires them but because change tends to create them for independent reasons; while some things that are rare or perhaps even nonexistent are not to be found because there are few if any pathways that could produce them from an available source".

[5] Interesting observations in this regard can also be found in Brown (forth).

can provide evidence for implicational universals, such as the generalizations on the order and the morphological marking of inflectional categories proposed by Greenberg (1963). And the Universals Archive[6] collects many cross-linguistic generalizations over form-function relations and asymmetries of this sort.

On the whole, cross-linguistic comparison of morphological forms has features as its major concern, a crucial and problematic area wherein typologists but also fieldworkers, computational linguists, and formal linguists tend to put increasing theoretical weight. In this book, the article by **Corbett** examines a number of aspects concerning the use of features and feature values both in individual language description and in typological comparison.

The discussion deals with two main questions: the *analysis problem* and the *correspondence problem*. The latter, which clearly reminds us of the more general problems of pre-established categories and their semantic comparability, is shown to have relevant consequences not only in cross-linguistic terms but also in intra-linguistic terms, for even within a single language, features and their values do not necessarily line up consistently. On the basis of well-selected examples, Corbett thus observes that greater care must be taken over these issues, identifying particular problems and solutions, but also showing the great advantages ensuing from such a scrupulous approach.

5 Universals and Syntax

The second area to which this book devotes three contributions is syntax, and this reflects the dominance of syntactic research in both Greenberg's most inspiring work and the last half-century linguistics in general.

Moreover, it is especially research on syntactic universals that has sparked off the debate between the functional-typological and the generative-formalist modes of explanation.

Croft's paper enters the debate by offering an excellent exposition of the crucial differences in methodology and argumentation. The author comes from the functionalist side, and his rigorous critique of the so-called "one language at a time" method, that is the tendency to focus in depth on a single language, and of the methodological opportunism adopted in many recent formal analyses, clearly points to the limits that the formalist mode of explanation inherited from structuralism.

The proposed five principles for a well-founded procedure in syntactic analysis and argumentation provide not only a valid defense from empirical disconfirmation, but also a sound endorsement of the functional-typological approach, and a major contribution to linguistic methodology as well.

Linguists working within generative frameworks, for instance, have taken a position where universals in syntax are usually held up as evidence for Universal Grammar, and where typology is often meant as a research method pursuing one

[6] Developed at the Universität Konstanz by F. Plank and his co-workers (http://typo.uni-konstanz.de/archive/intro/).

of the same goals as generative grammar: to contribute to a universal theory of grammar by determining the limits of possible human languages (Newmeyer 2005).

Considering cross-linguistic diversity and variation, however, functionalists take essentially the opposite stance: "[t]ypological theory is almost entirely unconcerned with distinguishing possible from impossible languages" (Nichols 2007: 232). And what's more, "typological theories are about probabilities of distributions, not about possibilities, and so they go far beyond the UG goal of defining what is possible" (Bickel 2007: 241).

As in phonology and morphology, distributions of syntactic patterns frequently reveal the effects of pervasive asymmetries. This subject is specifically examined in **Cinque**'s article that, from a methodological point of view, is a nice counterpart to Croft, since it argues not only for a formal approach to language universals, but also for the integration of typology and formal theory.

The author takes an incredibly rich amount of data from typological research in order to illustrate and explain the incidence of left-right asymmetries between functional modifiers and heads. Among the various left-right asymmetries of natural languages that have been repeatedly observed and discussed in the literature, the asymmetry at stake was already implicit in Greenberg's Universal 20, and is clearly instantiated by a number of different patterns in dozens of languages. The paper lucidly demonstrates that asymmetries cannot be handled in terms of consistent ordering of heads and dependents, but they can be derived by making assumptions about a single underlying word order and highly constrained possibilities for movement of elements within Kayne's Antisymmetry framework.

The role of heads in word order correlations was also discussed in Dryer's explanation for some Greenbergian universals in terms of branching versus nonbranching structures (Dryer 1992). In his contribution for this volume, **Dryer** proposes a refreshing and thought-provoking revision of the Branching Direction Theory. The discussion is honestly set out in light of recent work suggesting that many of these branchings, interpreted under the assumption of a flat phrase structure, are questionable. In some cases, the author shows that alternative explanations are available, such as processing constraints or the diachronic processes that gave rise to the constructions. Yet other examples, like auxiliary verb and main verb, do not seem explainable in this way and remain problematic: for these cases the need for further investigation is clearly admitted.

Despite claims that the Greenbergian approach and the Chomskyan approach to linguistic universals are diametrically opposed to each other in every respect, both are in fact examining the same phenomena, and these articles clearly suggest that sharing results in respective fields encourages the reflection on the methods and their refinement.

6 Universals and Semantics

Looking at semantic typology, Mairal and Gil say that "[i]n contrast to phonological and syntactic universals, very little attention has been paid to the study of semantic universals" (Mairal and Gil 2006: ix).

Actually, according to the most radical claim about linguistic relativity, the search of semantic universals may be doomed to failure; but this claim does not appear to be correct to a degree that would make cross-linguistic semantics impossible. And in point of fact, although differing in the degree to which semantic aspects of grammatical phenomena are foregrounded or backgrounded, most studies in grammatical typology do imply some kind of semantic orientation. Therefore, semantic typology considerably overlaps with lexical typology, here meant in its broadest sense, that is as concerned with typologically relevant patterns of the lexicon-grammar interaction (Koptjevskaja-Tamm et al. 2007: 160),

A number of studies and projects has effectively contributed to this search.[7] Apart from the proposals about universal inventories of lexical items, from Swadesh lists to the 'semantic primes' enumerated within the framework of Natural Semantic Metalanguage (Goddard and Wierzbicka 1994, Goddard 2001),[8] most of the 142 semantic universals in Plank's Universals Archive also describe interesting constraints on the lexicon. Moreover, systematic cross-linguistic research on semantic categorization has recently explored diverse conceptual domains typically encoded by words, such as *color, body, kinship, perception, motion, events of breaking and cutting, dimension*.

The study of categorization within lexical fields or conceptual domains is obviously confronted with diverse focal question, including polysemy, which is also relevant to cross-linguistic comparison of grammatical forms.

In this respect, just to mention a few choices, implicational scales and the method of semantic maps (Haspelmath 2003), which visually represent cross-linguistic regularity in semantic structure, have enjoyed increased popularity as empirically testable and reliable tools for expressing similarities of categories across languages.

The search for typologically significant correlations between lexicon and grammar also involves the problem of cross-linguistic variation in word classes,[9] whose behavior and classification can be better understood by taking into account lexical semantics and more fine-grained lexical distinctions (e.g. Dixon and Aikhenvald 2004).

And lexicon-grammar interaction constantly provides novel challenges for future lexical-typological research in formal semantics too, which is primarily concerned with compositionality, i.e. how the meaning of a complex expression can be derived from the meanings of its parts. Interestingly, Talmy (2007) develops a well-

---

[7] For the literature, which is too extensive to be listed here, see Koptjevskaja-Tamm et al. (2007) and von Fintel and Matthewson (2007).

[8] The former ones obviously proposed not as claims for universal lexical status, but as tools for large-scale lexico-statistical and glotto-chronological investigations. The latter ones variously disputed even among the individual contributors to Goddard and Wierzbicka (1994).

[9] Semantic characterizations and classifications of the speech parts seem hardly survive close scrutiny (e.g. Croft's discussion on Baker (2003)), and more able to predict only general cross-linguistic tendencies, but a proposal concerning the use of universal exemplars (basic words like the equivalents of *person*, *thing* for nouns, or *do*, *happen* for verbs) is to be found in Wierzbicka et al. (2000).

articulated proposal that cuts across the divide between functionalists and formalists by integrating work on semantic decomposition and cross-linguistic comparison in verbal semantics.[10]

A comparable attitude can be found in the article of **Delfitto**, which contains many interesting points in the proposed discussions on determiners, on the lexical operations explaining the causative/unaccusative alternations, and on the *Aktionsart* distinctions. The author argues that a lexicon involves a selection of proper subparts of conceptual structure and originates two distinct forms of linguistic computation: a lexical computation and a syntactic computation. Thus, the first form of semantic computation is pre-syntactic, whereas the second one is post-syntactic. Accordingly, different semantic universals can be identified in each of these two distinct domains.

7 Universals and Diachrony

The diachronic typological approach to language universals is an essential facet of many of the functionally inclined typologists, and this goes back at least to Greenberg's (1965, 1978) early calls for diachrony in typology, and to Givón's (1979) strong argumentation in this direction.

For modern typology, understanding distributions of linguistic types and patterns as historically grown implies that current synchronic distributions are seen as the product of history and diachronic processes.

As a matter of fact, typology is in itself an *explanandum* rather than an *explanans*, and in this respect "there is no need to choose between synchronic and diachronic accounts: synchrony is what diachrony explains" (Moravcsik 2007: 39). Yet, probabilities of change and the principles behind them are a matter of ongoing debate. Broadly speaking, it is possible to identify two basic approaches to these issues, namely whether universal tendencies and geographical clusterings result from overall principles that govern the evolutionary trends, or from locally motivated diachronic paths, as assumed by much work within the framework of grammaticalization theory.[11]

Under both views, type transitions and type modifying language internal developments are privileged objects of inquiry, which also may cast some light on the relation between linguistic reconstruction and typology (Fisiak 1997), in a domain where linguistic comparison goes back to its roots.

---

[10] As an example, van Valin (2006: 177) claims that the *Aktionsart* distinctions derived from Vendlerian verb classes "are among the most important organizing principles of verbal systems in human languages". But this insight, formulated solely on data from English, has been falsified by Tatevosov's (2002) comparative study. What is truly universal within actionality may not be the classes themselves, but rather the smaller building blocks from which event structures are composed and which all languages make use of in constructing their *Aktionsart* distinctions (von Fintel and Matthewson 2007).

[11] In the enormous amount of literature on this subject, Bybee's research in particular has been widely admired (e.g. Bybee et al. 1988, 1994, and 2003).

As it is well known, the earliest documents in Indo-European languages attest to a typological stage where syntactic relations are marked in the extremely complex morphology of nominal and verbal paradigms, with attendant great freedom in word order. The subsequent typological shift results in the appearance of the so-called configurational type, where the complex morphological paradigms are strongly reduced and the word order is not free.

The long-term evolutionary processes of word order in Latin are specifically investigated by **Magni**, whose article focuses on the gradual construction-by-construction reorientation of constituent order as reflected in coexisting patterns such as adpositions, coordinative, adjectival, genitival and comparative constructions. The author shows that many of the more far-reaching claims regarding consistency of word order type and diachronic change simply do not hold up to closer scrutiny. Ample evidence taken from documents ranging from early to classical Latin, and also comparison with Italic Languages and other Indo-European languages prove that the key features of Latin word order are more amenable to approaches like information structure.

8 Universals and Second Language Acquisition

The diachronic dimension in the study of human language also involves the peculiar developments whereby learner varieties are gradually built up through the various transitional stages of the interlanguages.

Not surprisingly, Greenberg himself proposed some important remarks concerning the relation between typology/universals and research on second language acquisition, explicitly claiming that: "universals apply equally to interlanguage and to primary language" (Greenberg 1991: 39).

In consonance with this, the congress held in Bologna also included the contribution of a leading scholar in both typological and SLA research: in her article **Giacalone** follows up on Greenberg's suggestions, especially focusing on the domain of morphology and paying particular attention to the complex area of tense/aspect/*Aktionsart*. In her discussion, the author makes use of the ample database available on the acquisition of Italian as a second language, and clearly points out to the advances in the fruitful interaction of typology and SLA studies, where more recent research programs have clearly demonstrated that L2 materials and data can not only contribute to explain how people acquire languages, but also to our understanding of how the human linguistic capacity is structured and functions. Such studies can provide useful insights and help to clarify theoretical issues, thus effectively contributing to research into universals of language.

9 Conclusions

Four decades of research have provided a more accurate body of knowledge about typologically relevant patterns, more reliable tools for their analysis, and substantial

advances in our understanding of universal tendencies and principles across the languages of the world. Major breakthroughs have taken place in both theoretical and empirical issues, with contributions by a number of linguists both working within the functional-typological framework and adhering to generative-formal positions.

The papers that make up this volume compare these different, but not irreconcilable, approaches to explaining linguistic universals, with a focus on up-to-date research dealing mainly with synchronic, but also diachronic data, and covering various areas of linguistics, as well as a wide range of languages.

In the overall picture emerging from these contributions, current research on language universals shows an almost baroque fascination with both the ever-attracting macrocosms of large-scale architectures, and the kaleidoscopic microcosms of fine-grained features and details. And both perspectives point to exciting areas for future research.

Elisabetta Magni
Sergio Scalise
Antonietta Bisetto

References

Aikhenvald, Alexandra. 2007. Typological distinctions in word-formation. In *Language typology and syntactic description*. Volume 3: *Grammatical categories and the lexicon*, ed. Timothy Shopen, 1–65. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2005. Morphological universals and diachrony. In *Morphology and linguistic typology. On-line Proceedings of the Fourth Mediterranean Morphology Meeting MMM4, Catania 21–23 September 2003*, eds. Geert Booij, Emiliano Guevara, Angeliki Ralli, Salvatore Sgroi, and Sergio Scalise, 45–57. University of Bologna. http://morbo.lingue.unibo.it/mmm/.

Baerman, Matthew, and Greville G. Corbett. 2007. Linguistic typology: morphology. *Linguistic Typology* 11: 115–117.

Baker, Mark C. 2003. *Lexical categories: Verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Bickel, Balthasar. 2005. Typology in the 21st century: Major current developments. *Linguistic Typology* 11: 239–251.

Bickel, Balthasar, and Joanna Nichols. 2007. Inflectional Morphology. In *Language typology and syntactic description*. Volume 3: *Grammatical categories and the lexicon*, ed. Timothy Shopen, 169–240. Cambridge: Cambridge University Press.

Brown, Dunstan. forth. Morphological typology. To appear in *The Oxford handbook of linguistic typology*, ed. Jae Jung Song. Oxford: Oxford University Press. http://www.surrey.ac.uk/LIS/SMG/dbpubs.htm.

Bybee, Joan L. 1988. The diachronic dimension in explanation. In *Explaining language universals*, ed. John A. Hawkins, 350–379. Oxford: Basil Blackwell.

Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994. *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.

Bybee, Joan L. 2003. Mechanisms of change in grammaticization: The role of repetition. In *Handbook of historical linguistics*, eds. Brian D. Joseph and Richard D. Janda, 602–623. Oxford: Basil Blackwell.

Coopmans, Peter. 1985. Language types: Continua or parameters? Ph.D. thesis. Utrecht: Rijksuniversiteit.

Cristofaro, Sonia. forth. Language universals and linguistic knowledge. To appear in *The Oxford handbook of linguistic typology*, ed. Jae Jung Song. Oxford: Oxford University Press.
Croft, William. 2000. Parts of speech as language universals and as language-particular categories. In *Approaches to the typology of word classes*, eds. Petra M. Vogel, and Bernard Comrie, 65–102. Berlin: Mouton de Gruyter.
Croft, William. 2001. *Radical construction grammar*. Oxford: Oxford University Press.
Croft, William. 2003. *Typology and universals*, 2nd ed. Cambridge: Cambridge University Press.
Darnell, Michael, Edith Moravcsik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley, eds. 1999. *Functionalism and formalism in linguistics*, 2 vols. Amsterdam/Philadelphia: John Benjamins.
Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.
Dryer, Matthew S. 1997. Are grammatical relations universal? In *Essays on language function and language type*, eds. Joan L. Bybee, John Haiman, and Sandra A. Thompson, 115–143. Amsterdam/Philadelphia: John Benjamins.
Dixon, R. M. W., and Alexandra Aikhenvald, eds. 2004. *Adjective classes*. Oxford: Oxford University Press.
Fischer, Olga. 2007. *Morphosyntactic change*. Oxford: Oxford University Press.
Fisiak, Jacek, ed. 1997. *Linguistic reconstruction and typology*. Berlin and New York: Mouton de Gruyter.
Giorgi, Alessandra, and Giuseppe Longobardi. 1991. *The syntax of noun phrases.* Cambridge: Cambridge University Press.
Givón, Talmy 1979. *On understanding grammar*. New York: Academic Press.
Goddard, Cliff. 2001. Lexico-semantic universals: A critical overview. *Linguistic Typology* 5: 1–65.
Goddard, Cliff, and Anna Wierzbicka, eds. 1994. *Semantic and lexical universals*. Amsterdam/Philadelphia: John Benjamins.
Graffi, Giorgio. 1980. Universali di Greenberg e Grammatica Generativa, *Lingua e Stile* 15: 371–387.
Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of grammar*, ed. Joseph H. Greenberg, 73–113. Cambridge, Mass: MIT Press.
Greenberg, Joseph H. 1965. Synchronic and diachronic universals in phonology. *Language* 42: 508–517.
Greenberg, Joseph H. 1978. Diachrony, synchrony and language universals. In *Universals of human language*, Volume 1: *Method and theory*, eds. Joseph H. Greenberg, Charles A. Ferguson, and Edith A. Moravcsik, 61–92. Stanford: Stanford University Press.
Greenberg, Joseph H. 1991. Typology/universals and second language acquisition. In *Crosscurrents in second language acquisition and linguistic theories*, eds. Thom Huebner, and Charles A. Ferguson, 37–43. Amsterdam/Philadelphia: John Benjamins.
Haspelmath, Martin 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
Haspelmath, Martin 2003. The geometry of grammatical meaning: Semantic maps and crosslinguistic comparison. In *The new psychology of language*, Volume 2, ed. Michael Tomasello 211–242. Mahwah, NJ: Erlbaum.
Haspelmath, Martin 2006. Parametric versus functional explanations of syntactic universals. To appear in *The limits of syntactic variation*, ed. Theresa Biberauer. Amsterdam/Philadelphia: John Benjamins. http://email.eva.mpg.de/~haspelmt/publist.html.
Haspelmath, Martin 2007. Pre-established categories don't exist. Consequences for language description and typology. *Linguistic Typology* 11: 119–132.
Haspelmath, Martin, Matthew S. Dryer, David Gil, and Bernard Comrie, eds. 2008. *The world atlas of language structures online*. Munich: Max Planck Digital Library. http://wals.info/index.
Hawkins, John A. 1983.*Word order universals*. New York: Academic Press.
Kayne, Richard. 1994. *The antisymmetry of syntax.* Cambridge (MA): MIT Press.
Koopman, Hilda. 1983. *The syntax of verbs*. Dordrecht: Foris.
Koptjevskaja-Tamm, Maria, Martine Vanhove, and Peter Koch. 2007. Typological approaches to lexical semantics. *Linguistic Typology* 11: 159–185.

Lehmann, Winfred P. 1973. A structural principle of language and its implications. *Language* 49: 47–66.

Lehmann, Winfred P. 1978. The great underlying ground-plans. In *Syntactic typology: Studies in the phenomenology of language*, ed. Winfred P. Lehmann, 3–55. Austin: University of Texas Press.

Mairal, Ricardo, and Juana Gil, eds. 2006. *Linguistic Universals*. Cambridge: Cambridge University Press.

Moravcsik, Edith A. 2007. What is universal about typology? *Linguistic Typology* 11: 27–41.

Newmeyer, Frederick J. 2005. *Possible and probable languages: A generative perspective on linguistic typology*. Oxford: Oxford University Press.

Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11: 133–157.

Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: The University of Chicago Press.

Nichols, Johanna. 2007. What, if anything, is typology? *Linguistic Typology* 11: 231–238.

Plank, Frans. 2007. Extent and limits of linguistic diversity as the remit of typology – but through constraints on WHAT is diversity limited? *Linguistic Typology* 11: 43–68.

Schachter, Paul, and Timothy Shopen. 2007. Parts-of-speech systems. In *Language typology and syntactic description*. Volume 1: *Clause structure*, ed. Timothy Shopen, 1–60. Cambridge: Cambridge University Press.

Siewierska, Anna. 1996. Word order type and alignment type. *Sprachtypologie und Universalienforschung* 49.2: 149–176.

Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.

Stowell, Tim. 1981. *Origins of phrase structure*. Ph.D thesis. Cambridge, MA: MIT Press.

Talmy, Leonard. 2007. Lexical typologies. In *Language tipology and syntactic description*. Volume 3: *Grammatical categories and the lexicon*, ed. Timothy Shopen, 66–168. Cambridge: Cambridge University Press.

Tatevosov, Sergej. 2002. The parameter of actionality. *Linguistic Typology* 6: 317–401.

Travis, Lisa. 1984. *Parameters and effects of word order variation*. Ph.D. thesis. Cambridge, MA: MIT Press.

van Valin, Robert D. Jr. 2006. Some universals of verb semantics. In *Linguistic universals*, eds. Ricardo Mairal, and Juana Gil, 155–178. Cambridge: Cambridge University Press.

von Fintel Kai, and Lisa Matthewson. 2007. Universals in semantics. To appear in *The linguistic review*, special issue devoted to universals. http://semantics-online.org/2007/07/universals-in-semantics.

Wierzbicka, Anna. 2000. Lexical prototypes as a universal basis for cross-linguistic identification of parts of speech. In *Approaches to the typology of word classes*, eds. Petra M. Vogel, and Bernard Comrie, 285–317. Berlin: Mouton de Gruyter.

How Universal are Linguistic Categories?

Paolo Ramat

Abstract From the arbitrary (though sensible) definition of language as "(phonic) system used to say something about someone or something" it follows that to say (i.e. to predicate) something about someone or something (i.e. about entities or states of affairs conceived in our mind) belongs to the basic activities of our brain. In Edward Sapir's words "There must be something to talk about and something must be said about this subject once it is selected" (Sapir 1921; repr. 1949: p. 119).

If we understand the Aristotelian terms *hypokeímenon* and *katēgoroúmenon* (Lat. *subjectum* and *praedicatum* respectively) not in the grammatical meaning they acquired in the Western grammatical tradition but, in a functional sentence perspective, as 'topic' and 'comment', or 'theme' and 'rheme', we may affirm that they constitute the basic sentence structure, the essential part for the semantic interpretation of the sentence. Consequently, many linguists see NOUN and VERB as universal categories that all languages must have. But the discussion concerning whether the distinction NOUN/VERB is valid everywhere, for instance among the Iroquoian languages, is far from being settled. At the other end of the extant typological structures, the same could be said for the so-called precategorial languages of South East Asia, in which the functional value of a word (and hence its categorial status) is often determined only by its syntactic context (see Walter Bisang's many contributions on the subject).

We need multiple criteria in order to assign a category to a lexeme, or better, a given lexeme to a category. Accordingly, the present paper tries to make a distinction between semantic function and morphosyntactic functioning of words and shows that both viewpoints are necessary and complementary to define the linguistic status of a word. The concept of 'tertium comparationis' will help to clarify the point.

Keywords Cognitivism · deduction · formalism · functionalism · induction · prototype · relativism

P. Ramat
University of Pavia

S. Scalise et al. (eds.), *Universals of Language Today*, 1–11,
DOI: 10.1007/978-1-4020-8825-4_1,

> 'without categorization, memory is useless [...] Thus an account of the organism's ability to categorize transcends linguistic theory. It is central to all cognitive psychology', Ray Jackendoff, *Semantics and Cognition,* 1985:77.

In order to give a historical dimension to the question included in the title of the present paper, it may be useful to start by a quote from the Latin grammarian Priscian (5th–6th cent. ; cp. Vineis 1998): *Partes igitur orationis* [i.e. of the sentence] *sunt, secundum dialecticos duae, nomen et verbum: quia hae solae, etiam per se coniunctae, plenam faciunt orationem: alias autem partes, syncategoremata, hoc hest consignificantia, appellabant* (Prisciani Institutio de arte grammatica, lib. II, 2). "The parts of the sentence, according the 'dialectici' are two: the noun and the verb, since only these [parts], even when joined in isolation [i.e. without any other part: P.R.] produce a meaningful sentence. The other parts were called 'syncategoremata', [which means that] they have meaning together with nouns and verbs". '*Plena oratio*' means meaningful sentence, self-sufficient utterance. The classical tradition often did not distinguish between form and function: see for instance the double meaning of *rhêma* as "verb" and as "predicate", or *ónoma* as "noun" and "subject" (cp. Ramat 2005: 87f. ; see also below). This classical tradition went through the centuries as testified, for instance, by the following dialogue fragment from Horne Tooke's famous *Diversions of Purley* (1786, quoted by Aarts 2006: 367f.):

> Beadon: *Well. For the present confine yourself to the* necessary *Parts: and exemplify in the English.*
> Tooke: *In English, and in all Languages* [my emphasis], *there are only* two *sorts of words which are* necessary *for the communication of thoughts.*
> Beadon: *And they are?*
> Tooke: *1. Noun and 2. Verb*

It can be seen that in Tooke's view 'noun' and 'verb' are basically considered from a functional rather than a syntactic viewpoint; in other words, more as topic and comment, theme and rheme, rather than as grammatical categories.

That NOUN and VERB (hereafter N and VB) are universal categories is maintained even in current discussions. I will limit myself to quoting Chomsky's paper on nominalization (Chomsky 1970), where N and VB are implicitly considered to be the essential syntactic features on the basis of which even the other major categories can be defined: ADJ is [+N,+VB], PREP is [-N, -VB] and, of course, N is [+N,-VB], while VB is [-N,+VB].

Conversely, we find the relativistic position expressed among others by Benjamin Lee Whorf in a well-known passage that deals precisely with substantives and verbs:

> "The Indo-European languages and many others give great prominence to a type of sentence having two parts, each part built around a class of word –substantives and verbs– which those languages treat differently in grammar [...] this distinction is not drawn from nature; it is just a result of the fact that every tongue must have some kind of structure, and those tongues have made a go of exploiting this kind. The Greeks, and especially Aristotle, built

> up this contrast and made it a law of reason [. . .] Our Indian languages [Hopi, etc.] show that with a suitable grammar we may have intelligent sentences that cannot be broken into subjects and predicates. Any attempted breakup is a breakup of some English translation or paraphrase of the sentence, not of the Indian sentence itself", Whorf 1956:241.

The discussion between universal formalism and relativistic functionalism has been taken up again in very recent times, in the frame of cognitive linguistics on the one hand, and linguistic typology on the other. Cognitive linguistics is concerned with how grammatical strategies are tied to more general cognitive abilities and more general cognitive processes (of categorization), thus making 'ad hoc' and redundant the generativists' hypothesis of a Universal Grammar (see, among others, Tomlin 1997:164; Tomasello 2003 and 2004, and the thorough discussion by Newmeyer on possible and probable languages: Newmeyer 2005).

However, this is not the place to reopen the long-lasting debate between relativism and universalism in language theories. As I shall conclude at the end of the present paper, I believe that the two positions are not irreconcilable: it is possible to find a common ground of agreement if we accept a prototypical approach to linguistic problems. In fact, Greenberg's famous probabilistic phrasing, which is also the general title of this Conference, "with more than chance frequency", hints at the possibility of inductively drawing universal conclusions from the observation of real language facts ('faits de langue').

Thus, let us go back to the linguistic discussion 'stricto sensu' alluded to at the beginning of this paper. Typology has shown that the functions of subjects (in the sense of AGENTs) and PREDICATEs may be expressed by very different strategies in different languages, while to say (i.e. predicate) something about someone or something (i.e. about entities or states of affairs conceived in our mind) does in fact belong to the basic activities of our brain. One of the possible definitions of language is indeed: "language is a (phonic) system used to say something about someone or something" or, in Edward Sapir's words, "There must be something to talk about and something must be said about this subject once it is selected" (Sapir 1921; repr. 1949: 119; see Gaeta 2002:20). If we understand the Aristotelian terms *hypokeímenon* and *katēgoroúmenon* (Lat. *subjectum* and *praedicatum* respectively) not in the grammatical meaning they acquired in the Western grammatical tradition but, in a functional sentence perspective, as 'topic' and 'comment', or 'theme' and 'rheme', we may affirm that they do constitute the universal basic sentence structure, the essential part for the semantic interpretation of the sentence.

Having distinguished between semantic functions and morphological forms, we may accept that the same semantic function be implemented by different morphs. Thus, we should not inquire whether theme and rheme, topic and comment are universal notions, but rather discuss whether grammatical categories such as NOUN, VERB, ADJECTIVE etc. may be considered language universals.

A comparison with the situation we find in phonology may help understand better what the question under discussion is. At the phonological level we know that in a vowel system with three vowels the most generally diffused opposition is /a/ ∼ /i/ ∼ /u/, whilst in systems with only two vowels the generally diffused opposition is between a low central vowel and a high one (as in Yimas, Papua New Guinea: see

WALS : Maddieson 2005:14). Exceptions to such generalisations are always possible. But at the phonological level we are dealing with physical objects (sounds) which can be accurately described in acoustic and articulatory terms. In Italian the [j] of *Ionio, iodio* is phonologically and phonetically different from the [i] of *piccolo, bambino* etc., though nowadays they are written with the same grapheme <i> (even if until the 19th century it was usual to write *noja, bujo* and even *studj*, plural of *studi-o*: *studi-i*). However, an [i] is an [i], i.e. a high front unrounded vowel in all languages of the world where it exists, whereas a [j] is a [j], i.e. a high front unrounded semivowel in all languages of the world where it exists (were this not true, the IPA would be useless!). Accordingly, we may conclude that [i] and [j] are sounds which can be realized in every human language, even if an implicational relation of the type [j] $\supset$ [i] could be affirmed, and [j] may be just an allophonic variant of /i/ (in today's Italian there are no minimal pairs based on the opposition /i/ $\sim$ /j/).

Things are different when we deal with notions such as NOUN, VERB, ADJECTIVE etc. (which belong to the theory of grammar and its implementations). One cannot have the physical perception of a noun in the same way one physically perceives an [i]. Concepts such as Ns, VBs, ADJs are not physical objects, although they are physically realised by sounds: they belong to a different level.[1] We can tentatively define an ADJ as a lexeme which (among other properties) assigns a quality to a referent and therefore modifies a head noun (: *the/a white house*; Ramat 2005:77). But this property is also shared by participles, which belong to the verbal paradigm:

(1) *La notte passata non sono andato a letto* "Last night I did not go to bed"
(2) *le notti passate non ho chiuso occhio* "During the last nights I did not sleep a wink".

Notoriously, Japanese ADJs ending with $-i$ have been considered as VBs:

(3) *ano tatemono wa takai*
this building TPC high i.e. "this building is high"

According to Bhat (1994:199) "Japanese [...] does not belong to the group of languages in which adjectives form a distinct category". This point has been discussed and refuted by Lombardi Vallauri (2000) who assigns both types of ADJs –the *-i*-ADJs as well as the *-na*-ADJs[2]– to a class to be distinguished both from VBs and Ns, according to a series of parametric features which also include the

[1] This holds also for the semantic level where there have been attempts to find 'semantic primes', i.e. universal notions which cannot be defined in terms of higher primitive concepts and are expressed in every language. Suffice to remember Anna Wierzbicka's studies on 'Natural Semantic Metalanguage' (e.g. Wierzbicka 1996; see also Goddard 2001), but this is not the place to enter the discussion on the validity of semantic universals. It must be noted, however, that recent studies of body parts' nouns, which one would be inclined to consider as best candidates to universality, have shown that they are in fact subject to different linguistic categorisations and language-specific principles: see *Language Sciences* 28 /2006, special issue "Parts of the body: cross-linguistic categorisation", ed. by N.J. Enfield, A. Majid and M. van Staden. (with Wierzbicka's reply: 2007).
[2] For example, *bakana hito* "stupid person".

linear order in the syntax of the sentence. The *-i*-ADJs partially overlap with VBs, the *-na*-ADJs with Ns, and both *-i*-ADJs and *-na*-ADJs share some morphological, syntactic and semantic properties, which allow us to speak of a class on its own. Both ADJs express properties, whilst VBs designate processes, activities or states and Ns designate things, mental or physical entities (Lombardi Vallauri 2000:333).

All in all, the Japanese situation is not qualitatively different from what can be observed for participles.[3] Participles behave like ADJs on the one hand (ex.s (1) and (2) above) but, on the other hand, they may have a clearly verbal function as in

(4) *La notte è passata* (**passato* !) "the night is over",
(5) *Ha passato/(?-a) tutta la notte senza dormire* "(s)he has not slept all night long"

Even without an AUX the participle may keep its verbal value:

(6) *Passata la notte, smise di piovere* "When the night was over it stopped raining".

Notice the difference between (6) and (1), i.e. between the verbal and the adjectival value of the very same form respectively.

Thus, semantics plays an important role even in assigning morphological roles. Consider compounds such as *(a/the) fireplace* or *(a/the) fireman.* The heads of the compounds are *place* and *man* respectively–exactly as is *house* in the NP *(a/the) white house* and in those forms which became proper names, written as a single word, such as *Whitehall*, *Whitehorse,* etc. It follows that in order not to consider *fire* as an ADJ (on the basis of its constructional parallelism with *Whitehorse*) we have to take account of formal as well as semantic-functional criteria. A *fireplace* is not a fire but 'a place for fire'.

Formal criteria may help to make the correct category decision. Even limiting ourselves to English, where ADJs do not show agreement with their head nouns, from a morphological point of view *fire* cannot have a comparative nor a superlative degree, which is a property of prototypical ADJs (cp. *whiter*, *whitest*: see below). Moreover, we can have *fires* (Plur.), but not **whites.* In its turn, from a syntactic point of view, *white* cannot be the head of an NP, unless it is substantivized via a syntactic device such as the insertion of a DET:

(7) *The brilliant white of the flowers.*

It is well-known that from the point of view of content we may have nominal constructs which have a verbal function and vice versa. Giuliana Fiorentino (2004)

---

[3] Indeed, the name the ancient grammarians gave to this particular form of the verbal paradigm, namely *metochḗ* "participation, sharing", calqued in Latin as *parti-cipium*, refers to its double function (on the mixed, non-prototypical character of the participles see Pompei 2006). This name contradicts the Aristotelian 'all-or-none' definition of *katēgoría*, according to which 'tertium non datur', as maintained by Frege ["The law of excluded middle is really just another form of the requirement that the concept should have a sharp boundary", Frege 1903 > 1997: 259, quoted in Aarts 2006:364]. In other words, an element (also a word) would belong to category A or to category B (cp. Ramat 2005: 87.). On the difference between categorial concepts and their implementation see below.

has correctly spoken of Action Noun Constructions as 'substitutive strategies' for verbal predicates, as in

(8a) *The destruction*$_{AN}$ *of the city by the enemy was terrible* versus
(8b) *The enemy destroyed the city in a terrible way*.[4]

On the other hand, it is well-known that also Infinitives may be replaced by Action Nouns (*nomina actionis*), especially in agglutinative languages, whereas fusional languages seem to prefer infinitive forms which belong more strictly to the verbal paradigm (cp. Sgall 2006: 410–417).[5]

Cross-linguistic data, as well as (neuro)psychological tests concerning linguistic production and comprehension, lead to the conclusion that the distinction between nouns and verbs is not neatly dichotomous but may be thought of as a continuum between two poles (cf. for a psychological-cognitivist approach Laudanna / Voghera 2006:xiii., with references. For a more linguistically oriented approach see, e.g., Simone 2003 and 2004 who has studied the Arabic *'al-maşdar* construction). Remember Benveniste's (1950>1966) statement that the difference between Ns and VBs has to be sought not in an opposition between 'things' and 'processes' but between two 'viewpoints' of the same state of affairs, in their variable form and their invariant referential function, as in (8a) and (8b).

When speaking of infinitives and action nouns we are on the verge between two different categories, as in the case of ADJs and participles. One might wonder why in Latin grammars the forms in *–bilis* such as *solubilis* "that can be solved, untied" or *laudabilis* "praiseworthy" are not included in the verbal paradigms of *solvere* and *laudare* whereas in Greek grammars the so-called 'verbal ADJs' (like *lytós* "untied, *solutus*" and "that can be untied, soluble", *lytéos* "that has to be untied, *solvendus*") are quoted among verbal forms. Where is the difference?

The discussion about what is a paradigm would lead us astray and too far from the main topic of this paper. What can be said from a general point of view is that the concept of prototype can be applied also to the notion of paradigm. On the basis of the examples quoted above, we see that there are forms which belong to the core of the verbal paradigm (say, the 1st person of the active indicative present: e.g. Lat. *laudo*) and more peripheral forms which may share some features with members of different categories (e.g. Latin gerundives and participles show an adjectival agreement with their head noun in gender, number and case). This is not detrimental to the fact that VB and ADJ build up two different categories, each of them prototypically characterized by a certain set of features.

---

[4] Note, however, that (8a) is not a complete sentence without *was terrible*, whereas (8b) is a complete sentence also without the specifying manner adverbial *in a terrible way.* This means that there are constraints on the use of different constructions and that, consequently, constructions have their own meaning.

[5] Constrast Turk. *gitmem lazım*, lit. 'my go is necessary' with its Engl. translation "I must go, I have to go".

As we shall see below, the categorial properties of a category (say, ADJ) belong to different levels of the linguistic organization (semantic-functional, morphosyntactic, pragmatic etc.; see Laudanna / Voghera 2006:xii).

In a previous article on the universality of grammatical categories (Ramat 1999, expanded in Ramat 2005: 68–73), I noted that a language may lack a given category and quoted from Schachter (1985: 15) the case of Hausa, in which very few ADJs exist. In order to say "a kind person" this language says:

(9a) *mutum* *mai* *alheri*
person having kindness;

or, in the predicative function:

(9b) *yana* *da* *alheri*
he.is with kindness : "he is kind".

Languages without any ADJ at all are Tuscarora, Chinese, Yurok (cp. Dixon 1977). These languages show a kind of morphological neutralization between the category ADJ and the category or categories which, so to speak, have taken over the adjectival function of expressing 'property concepts'. Hengeveld (1992) has proposed a scale of 'flexible' and 'rigid' languages with regard to the four major parts-of-speech, namely VB, N, ADJ, and ADV. Rigid languages are those which, like English, distinguish four parts-of-speech, or at least three of them like Dutch. Flexible languages are those which distinguish only two parts (e.g. Quechua, Turkish) or do not make any distinction between VB, N, ADJ and ADV, like Tongan. In these languages the syntactic as well as the semantic function of a lexeme is defined by its position within a construction.

In spite of these considerations, which seem to lead to an absolute relativity of categorial concepts, I believe that VB, N, ADJ etc. are universal categories and good tools of analysis for all languages.

The absence of the category ADJ in Tongan does not impinge on the existence of that category at the theoretical level. A quote from what Coseriu said on the occasion of the 11th International Congress of Linguists in Bologna may clarify what is meant here:

> . . . si l'on définit universellement un adjectif, ceci ne signifie aucunement que l'on attribue l'adjectif à toutes les langues, puisqu'une définition n'est pas un jugement d'existence, Coseriu 1974:49.

Furthermore, the implementation of the category ADJ may be different in different languages. I already mentioned languages which do not have the category at all. Also there are languages whose adjectives lack some of the prototypical adjectival features. Even inside the same language there may be more or less adjectival lexemes (as is the case with participles).

Hence clear-cut formal criteria are not sufficient to deal with (linguistic) categories and categorizing operations (see Wiemer / Bisang 2004: 7).

Categorization is a cognitive process of systematisation of what we experience and know (see the quote from Jackendoff at the beginning of this paper). Any time

we come across something new in our worlds, we want to assign it to some already known set of material entities or abstract concepts (cf. Aarts 2006:361). Therefore, it makes sense even for linguists to try and assign a linguistic element to a class of already known elements on the basis of shared properties or similarities.

Several kinds of criteria have been considered for the identification of linguistic categories. Firstly, categorial statements are often made on the basis of meaning. Words designating (material or mental) objects are labelled as Ns, whereas words referring to states of affairs (including actions, events) are labelled as VBs. Secondly, as we have already seen, the morphological structure also needs to be taken into account, although markers such as gender, number, cases may be shared by VBs, Ns, ADJs and even PROs, so that they cannot be taken as the only decisive parameter. Finally, one has also to consider the syntactic slots a lexeme may fill within a sentence. These three criteria may lead to different results even inside the same language. What might be identified as a verb on formal morphological grounds could serve, on the basis of its syntactic behaviour, the semantic function of a noun, namely to designate objects (that can be [± abstract]):

(10a) [[*il*$_{\text{DET}}$ *fumare*$_{\text{INF}}$]$_{\text{NP}}$ [≡ *il fumo*]$_{\text{NP}}$ *danneggia*$_{\text{TRANS.VB}}$ *gravemente la salute*] $_{\text{VP}}$,
lit. 'the to-smoke endangers heavily your health'

whereas in

(10b) [*fumare danneggia gravemente la salute*]$_{\text{VP}}$
fumare is a bare verbal form with the same SUBJ function (=AN) as *il fumare,* but syntactically cannot be labelled as an NP.

Tahitian *ao* may mean both "day" and "it's getting light", but in the sentence

(11) *'ua ao*

the aspectual morpheme *'ua* marks beyond any doubt *ao* as having a verbal function (cf. Lazard / Peltzer 2000:63f.).

Among isolating languages, it is well-known that the languages of (South)East Asia have lexical items which can be considered precategorial in the sense that they do not belong to any particular morphosyntactic category and show strong distributional versatility. It is the pragmatic context with its argument structure that disambiguates their function in discourse (see Bisang, forthcom.). This is not the norm in (old) Indo-European languages, which show a strong morphological array. In these languages in fact, difference of form entails category difference (e.g. *fun* vs. *funny*), though identity of form does not entail categorial identity (e.g. *clean* VB and ADJ). Bisang writes that "there are no obligatory grammatical categories in East and mainland Southeast Asian languages" and quotes the following Late Archaic Chinese example (Bisang 2006: 589)

(12) *wǒ*$_1$ *bú jiàn tā*$_2$ *yǐ̌ shì sān shíduō nián; jīntiān* ϕ_1 *jiàn* ϕ_2 *le*
I NEG see he already be 30 more year; today see PF
"I haven't seen him for more than 30 years. Today [I] saw [him]"

where the two arguments "I" (*wǒ*) and "he" (*tā*) are dropped in the second clause. Obviously, this is not meant to be a counterargument to the existence of the PRO category in Chinese. Needless to say that lexeme precategoriality or 'categorial flexibility', as Stéphane Robert (2004: 138) calls it, does not impinge on the existence of categories in general.

In conclusion, we have to distinguish between categories as theoretical concepts, which represent abstractions used by the linguist in his/her analysis, and practical, real implementations of these concepts. The former are universal and valid for the analysis of all languages (and indeed we speak of 'verbiness' or 'nouniness' also to account cross-linguistically for those languages which do not show a clear-cut distinction between VBs and Ns: cp. Ramat 2005:87). The latter may be language-specific or shared only by a certain number of languages.

In sum, between the absolute relativism of the Sapir-Whorf hypothesis and the assumption of a Universal Grammar there is, I think, a third way of approaching the problem.

The gradient character of linguistic categories with core representatives and more peripheral items is more and more recognized by many linguists. What Croft (1991) considers 'transitory categories', such as the auxiliary verbs, are just intermediate stages in the process of transcategorization (think for instance of ADJs used as ADVs such as Germ. *schnell*, It. *veloce* "swift" etc.: Ježek / Ramat, forthcom.). The final stage of a transcategorization process may represent the entering into a new morphosyntactic category as the final stage of a grammaticalization process (e.g. Engl. *during* originally an *–ing*-form of the VB *to dure*, which does not exist any longer; nowadays a PREP). Some forms may reach the endpoint of the transcategorization process, as is the case with participles becoming nouns (e.g. Hung. *költő* "poet", *elárusító* "shop-assistant" from the verbs *költ* "to create" and *elárusít* "to sell", It. *cantante* "singer", *calmante* "painkiller"); others do not, as is the case for *schnell* and *veloce*; others may even stop the transcatagorization process—see for instance Old It. *(u)omo* "man" used as indefinite pronoun just as French *on,* Germ. *man*. Nowadays, however, the indefinite use of *uomo* does not exist any longer. It has been replaced by the *si*-construct as in *si parla italiano* (vs. Fr. *on parle italien*, Germ. *man spricht Italienisch*) and *(u)omo* has kept just its nominal meaning (see Giacalone Ramat / Sansò, 2007). If we consider the diachronic dimension, there is no need for introducing new categories that would lie between N and PRO. We have simply to observe that recessivity is always possible. Along with developments N > N/PRO > PRO we may have also N> N/PRO > N.

Like Ockham (1258–1349) I deem that 'entia non sunt multiplicanda (praeter necessitatem)' "one must not multiply what exists (if not necessary)". Old, traditional categories, if considered both formally and functionally, are good enough to account cross-linguistically for parts-of-speech and their transformations.

I began this paper by quoting the Latin grammarian Priscian. Now I would like to conclude by referring to Thomas of Erfurt, the 'modista' of the 14th century, who considered pre-linguistic concepts to be an amorphous substance that acquires linguistic form only when it acquires a 'modus significandi' "a way of meaning", i.e. when it is assigned to a grammatical category.

Abbreviations

ADJ		PF	Perfect	
AN		PREP	Preposition	Adjective
AUX	Auxiliary	PRO	Pronoun	Action Noun
DET		TPC	Topic	
INF	Infinitive	VB	Verb	Determiner
N		VP	Verb Phrase	Noun
NP	Noun Phrase			

References

Aarts, Bas. 2006. Conceptions of categorization in the history of linguistics, *Language Sciences* 28: 361–385.

Bhat, Shankara D.N. 1994. *The adjectival category. Criteria for differentiation and identification.* Amsterdam/Philadelphia: Benjamins.

Benveniste, Émile. 1950 [>1966]. 'La phrase nominale', *Bull. de la Soc. de Linguist. de Paris* 46 [repr. in E.B., *Problèmes de linguistique génerale*. Gallimard, Paris, vol. I, Chapt. XIII].

Bisang, Walter. 2006. Southeast Asia as a linguistic category. In ed*Encyclopedia of language and linguistics*, 2nd edition, ed. -in-chief K. Brown, vol, 11, 587–595. Elsevier.

Bisang, Walter. forthcom. Transcategoriality and argument structure constructions in Late Archaic Chinese. In *The cognitive organization and reorganization of grammatical constructions* ed. Jaakko Leino. Amsterdam/Philadelphia: Benjamins.

Bisang, Walter, Nikolaus P. Himmelmann, Björn Wiemer. eds. 2004. *What makes grammaticalization? A look from its fringes and its components.* Berlin and New York: Mouton de Gruyter.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, eds. R. Jacobs and P. Rosenbaum, 184–221. Waltham, MA: Ginn & Co.

Coseriu, Eugenio. 1974. Les Universaux linguistiques (et les autres). In *Proceedings of the 11th International Congress of Linguists*, ed. L. Heilmann, vol. I: 47–73. Il Mulino, Bologna.

Croft, William. 1991. *Syntactic categories and grammatical relations: The cognitive organization of information.* University of Chicago Press: Chicago/London.

Dixon, Robert M.W. 1977. Where have all the adjectives gone? *Studies in Language* 1: 19–80.

Fiorentino, Giuliana. 2004. Nomi d'azione e subordinazione in italiano. *Studi e Saggi Linguistici* 42: 9–41.

Frajzyngier, Zygmunt, Adam Hodges, David S. Rood. eds. 2004. *Linguistic diversity and language theories.* Amsterdam/Philadelphia: Benjamins.

Frege, Gottlob. 1903/1997. *Grundgesetze der Arithmetik* . In *The frege reader*, ed. B. Michael. Blackwell, Oxford, vol. II.

Gaeta, Livio. 2002. *Quando i verbi compaiono come nomi. Un saggio di Morfologia Naturale.* FrancoAngeli, Milano.

Giacalone Ramat , Anna, Andrea Sansò. 2007. The spread and decline of indefinite *man*-constructions in European languages: An areal perspective'. In *Europe and the Mediterranean as linguistic areas: Convergencies from a historical and typological perspective*, eds. P. Ramat and E. Roma. Amsterdam/Philadelphia: Benjamins.

Goddard, Cliff. 2001. Lexico-semantics universals: a critical overview. *Linguistic Typology* 5: 1–65.

Hengeveld, Kees. 1992. Parts of Speech. In *Layered structure and reference in a functional perspective*, eds. M. Fortescue, P. Harder, L. Kristoffersen, 29–55. Amsterdam/Philadelphia: Benjamins.

Ježek, Elisabetta and Paolo Ramat. forthcom. On Parts-of-Speech transcategorization. *Folia Linguistica*.

Laudanna, Alessandro and Miriam Voghera. (a c. di). 2006. *Il linguaggio*. Strutture linguistiche e processi cognitivi. Bari/Roma: Laterza.

Lazard, Gilbert and Louise Peltzer. (2000). *Structure de la langue tahitienne*. Leuven/Paris: Peeters.

Lombardi Vallauri, Edoardo. 2000. Gli aggettivi giapponesi fra nome e verbo. *Studi Italiani di Linguistica Teorica e Applicata* 29: 311–345.

Maddieson, Jan. 2005. Vowel quality inventories. In *World atlas of language structures*, eds. M. Dryer, M. Haspelmath, D. Gil, B. Comrie, 14–15. Oxford: University Press.

Newmeyer, Frederick J. (2005). *Possible and probable languages: A generative perspective on linguistic typology.* Oxford: University Press.

Nuyts, Jan and Eric Pederson. eds. 1997. *Language and conceptualization*. Cambridge: University Press.

Pompei, Anna. 2006. Participles as a non-prototypical word class. In *Word classes and related topics in Ancient Greek*, eds. E. Crespo, J. De la Villa, and A. R. Revuelta, 361–388. Proceedings of the Conference held in Madrid, June 2003. Louvain-la-Neuve: Peeters.

Ramat, Paolo. 1999. Linguistic categories and linguists' categorizations. *Language* 37:217–251.

Ramat, Paolo. 2005. *Pagine linguistiche.* Bari/Roma: Laterza.

Robert, Stéphane. 2004. The challenge of polygrammaticalization for linguistic theory. In ed. Z. Frajzyngier et al. :119–142.

Sapir, Edward. 1921. *Language*. New York: Harcourt, Brace & World. [Repr. 1949].

Schachter, Paul. (1985). Parts-of-speech systems. In ed. T. Shopen, vol. I, 3–61.

Sgall, Petr. ed. 2006 [< 1958]. Zur Typologie des Infinitivs. In *Language in its multifarious aspects*. Prague: The Karolinum Press.

Shopen, Timothy, ed. 1985. *Language typology and syntactic description.* 3 vols. Cambridge: Cambridge University Press.

Simone, Raffaele. 2003. Maşdar, 'ismu al-marrati et la frontière verbe/nom. In *Estudios ofrecidos al prof. J.J. de Bustos Tovar*, eds. Girón Alconchel and José Luis et al., 2 vols, 259–288. Madrid, Editorial Complutense.

Simone, Raffaele. 2004. L'infinito nominale nel discorso. In *Generi, architetture e forme testuali*, vol. I: 73–96. Atti del VII Convegno SILFI, Roma ottobre 2002. Roma, Cesati.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition.* Cambridge (Mass.): Harvard University Press.

Tomasello, Michael. 2004. What kind of evidence could refute the UG hypothesis? Comments on Wunderlich. *Studies in Language* 28: 642–645.

Tomlin, Russell S. 1997. Mapping conceptual representations into linguistic representations: the role of attention in grammar. In eds. J. Nuyts and E. Pederson, 162–189.

Vineis, Edoardo. 1998. *Partes orationis*: parti del discorso o parti della proposizione? In *Ars linguistica*, a c. di. G. Bernini, P. Cuzzolin, and P. Molinelli, 521–526. Studi offerti a Paolo Ramat. Roma: Bulzoni.

Whorf, Benjamin Lee. 1956. *Language, thought, and reality* ed. by J. B. Carroll, Cambridge (Mass.).

Wierzbicka, Anna. 1996. *Semantics: Primes and universals*. Oxford: University Press.

Wierzbicka, Anna. 2007. Bodies and their parts: An NSM approach to semantic typology. *Studies in Language* 29: 14–65.

An Empirical Test of the Agglutination Hypothesis[1]

Martin Haspelmath

Abstract In this paper, I approach the agglutination-fusion distinction from an empirical point of view. Although the well-known morphological typology of languages (isolating, agglutinating, flexive/fusional, incorporating) has often been criticized as empty, the old idea that there are (predominantly) agglutinating and (predominantly) fusional languages in fact makes two implicit predictions. First, agglutination/fusion is characteristic of whole languages rather than individual constructions; second, the various components of agglutination/fusion correlate with each other. The (unstated, but widely assumed) Agglutination Hypothesis can thus be formulated as follows:

(i) First prediction: If a language is agglutinating/fusional in one area of its morphology (e.g. in nouns, or in the future tense), it shows the same type elsewhere.
(ii) Second prediction: If a language is agglutinating/fusional with respect to one of the three agglutination parameters (a-c) (and perhaps others), it shows the same type with respect to the other two parameters: (a) separation/cumulation, (b) morpheme invariance/morpheme variability, (c) affix uniformity/affix suppletion.

I report on a study of the nominal and verbal inflectional morphology of a reasonably balanced world-wide sample of 30 languages, applying a variety of measures for the agglutination parameters and determining whether they are cross-linguistically significant. The results do not confirm the validity of the Agglutination Hypothesis, and the current evidence suggests that "agglutination" is just one way of trying to capture the strangeness of non-Indo-European languages, which all look alike to Eurocentric eyes.

M. Haspelmath
Max-Planck-Institut für evolutionäre Anthropologie, Leipzig

[1]Earlier versions of this paper were presented at the 3rd conference of the Association for Linguistic Typology (Amsterdam 1999) and at the 9th International Morphology Meeting (Vienna 2000). I am grateful to the audiences at these occasions, as well as to two reviewers, for useful comments. (The core ideas of this paper were first presented at the DGfS Summer School on Language Typology in Mainz, September 1998.)

S. Scalise et al. (eds.), *Universals of Language Today*, 13–29,
DOI: 10.1007/978-1-4020-8825-4_2,

Keywords Typology · typological morphology · inflection · agglutination · fusion

1 Agglutination and fusion: An Ambiguous Success Story

One of the seemigly most successful stories in the history of linguistic typology is the creation of a holistic morphological typology in the first half of the 19th century, initially by the combined efforts of Friedrich von Schlegel (1808), August Wilhelm von Schlegel (1818), and Wilhelm von Humboldt (1822, 1836). As is widely known (e.g. Greenberg 1974: 35–41), these three were responsible for the classical subdivision of languages into an **isolating** (or "analytic") type, an **agglutinating** type, a **fusional** (or **flexive**) type,[2] and an **incorporating** type. This way of classifying languages was made popular especially by Schleicher (1850) and Müller (1871), and has been part of linguists' textbook knowledge ever since. Almost every introduction to linguistics mentions the terms, and they are frequently used in the technical literature (at least the term *agglutinating/agglutination*). Three representative sentences from recent works by influential authors are given in (1).

(1) a. Evans (1995: 1): "Kayardild is a dependent-marking, agglutinating, entirely suffixing language with a free order of phrasal constituents and a rich system of case-marking. . ."
b. Slobin (1997: 281): ". . . On this account, agglutinating languages like Turkish and Japanese have no closed-class morphemes. . ."
c. Hyman (2001: 1397): ". . . In other words, a highly developed paradigmatic system of tonal oppositions appears not to be very compatible with a highly developed syntagmatic system of agglutinative morphology"

But at least since Sapir (1921), it has been widely recognized that this four-way distinction is problematic, because it conflates three different parameters: (i) the **degree of synthesis** (isolating vs. non-isolating, or in other terms, analytic vs. synthetic); (ii) the **degree of stem combination** (incorporating vs. non-incorporating); and (iii) the **degree of fusion** (agglutinating vs. fusional/flexive). The notions of synthesis and stem combination are quite easy to describe and identify, as long as one agrees on word boundaries and a definition of "stem". But what exactly is meant by agglutination and fusion/flexion, and what we need such concepts for, cannot be so readily explained.

---

[2] For the flexive type, other term variants such as *(in)flexional/inflectional* are also often used. As was often noted (e.g. Bazell 1958), the term *inflectional* is confusing because it also has a different sense: One also says that agglutinating languages have inflection (i.e. different word-forms belonging to a single lexeme), so Sapir's term *fusional* has tended to supplant it in the typological sense. Plank (1999) retains the term *flexive* (deliberately differentiating it from *inflectional*), presumably because he feels that there is much more to the agglutination/flexion distinction than what Sapir meant by *fusion* (cf. note 7). In this paper, I use *fusion* as the opposite of *agglutination*, simply because it seems that this term is now better known.

So with respect to the agglutination/fusion distinction, the success of the 19th century classification is ambiguous: While the classification is still widely known, it does not have an exact meaning and does not seem to be taken seriously. In fact, with a few notable exceptions (Skalička 1951/1979, Plank 1986, 1991, 1999, Plungian 2001, Testelets 2001), linguistic theorists in the latter half of the 20th century have either ignored or severely criticized 19th century morphological typology. The quotations in (2) seem to be fairly characteristic of the mainstream view.

(2) a. Anderson (1985: 10): ". . . nothing much seems to follow from this classification: it has never been shown, for example, that languages with agglutinative properties share other features of a non-accidental sort that are not shared with non-agglutinating languages as well. For these reasons, the traditional terms do not seem to constitute any significant typology."
b. Bauer (1988: 170): "Basically, a typology is not of much value unless it predicts other things about the various types of languages. . . Now . . . a typology in terms of isolating, agglutinative and fusional does not [seem to] correlate with anything else in the morphology at all. . . The value of the typology *qua* typology is thus very much in doubt."
c. Spencer (1991: 38) "This typology, though sanctioned by tradition, has been criticized for being both incoherent and useless. It is useless because nothing of any interest follows from classifying languages in this way."

But if classical morphological typology is "incoherent and useless", why is the terminology still with us? Moreover, most linguists find that the parameters of degree of synthesis (e.g. Greenberg 1954[1960], Bickel & Nichols 2005) and incorporation (e.g. Baker 1996) are of great typological interest. Is there also an "agglutination parameter"? In the following section I will argue that the agglutination/fusion distinction is much more interesting than superficial textbook statements such as in (2a–c) make it appear, but we will eventually see that an empirical test largely comes to negative conclusions about ist validity.

2 The Agglutination Hypothesis: Implicit Claims Made Explicit

Anderson, Bauer and Spencer make seriously misleading statements when they claim that nothing of interest follows from classifying languages into agglutinating and fusional types, and that it implies no correlations. The use of this classification does imply testable claims about extremely interesting correlations, but these claims are usually not made explicit.

The reason why the classification implies a number of correlations is that "agglutinating language" is not a primitive notion. As we will see shortly, saying that a language or a pattern is agglutinating embodies a set of logically separate claims. By accepting the classification of languages or patterns into agglutinating and non-agglutinating types, one implicitly accepts the idea that the various properties that make a language agglutinating correlate with each other. Thus, we should think of agglutination not so much as a classificatory concept or as a parameter

of variation, but as an empirically testable hypothesis. I call it the **Agglutination Hypothesis** here, and I distinguish two main components of it. (This is very similar to Plank's (1999:285) "Strong Homogeneity Hypothesis".[3])

The first empirical claim (or universal prediction) follows from the fact that what is normally classified as agglutinating or fusional is whole languages, not just small subsystems. So implicit in this classification is the expectation that the First Prediction in (3) is correct.

(3) **First Prediction: Correlation among parts of the morphology**
If a language is agglutinating/fusional in one area of its morphology (e.g. in nouns, or in the future tense), it tends to show the same type elsewhere.

Logically speaking, languages could of course be agglutinating in their noun morphology and fusional in their verb morphology, or even isolating in the indicative mood and agglutinating in the subjunctive mood. But linguists tend to accept the notion that the morphological types apply to entire languages, thus implicitly claiming that languages are more homogeneous than they would have to be. This presupposition of course goes back to the early 19th century, and it could be that it is simply naive. In the early days of modern typology, what struck linguists was the differences between languages, and not so much the more abstract differences between patterns. In other areas of typology, linguists also tended to ascribe interesting differences to languages in the past (e.g. "ergative languages", "tone languages"), whereas more recently they have usually become more careful, recognizing that system splits (such as ergative in the perfective vs. accusative in the imperfective) are not a marginal phenomenon. Dividing languages into accusative, ergative and neutral languages is a meaningful enterprise only if one expects that different constructions tend to show the same patterns, or at least that one of the constructions is somehow predominant.

The second empirical claim follows from the fact that agglutination is not a primitive feature. But what exactly are the component features of agglutination? Unfortunately, it is very difficult to find a precise definition of agglutination and fusion in the literature. A few exemplary definitions are listed in (4) (emphasis is mine).

(4) a. Bloomfield (1933: 207): "*Inflectional* languages show a merging of semantically distinct features either in a **single bound form** or **in closely united bound forms**, as when the suffix *-ō* in a Latin form like *amō* 'I love' expresses the meanings 1st, sg, pres, ind."
b. Hjelmslev (1963[1968:109]): "... *flektierenden* Typ, bei dem **die Grenze zwischen Wurzel und Suffix nicht klar ist,** bei dem jedes Suffix oftmals gleichzeitig mehrere verschiedene grammatische Verhältnisse ausdrückt, und

[3] The only reason I do not adopt Plank's term is that my term "Agglutination Hypothesis" appears in the title of this paper, and I did not want the paper to be unrecognizable. Note that I developed these ideas independently of Plank (but under the influence of his earlier work, e.g. Plank 1986, 1991), at about the same time.

bei dem **die Wurzel selbst bei der Flexion Veränderungen erfahren** kann."[4]

c. Anderson (1985: 9): "*(in)flectional* languages have internally complex words which **cannot easily be segmented** into an exhaustive and non-overlapping string of formatives"

d. Vance (1987: 175): "[In agglutinative languages] **each morph represents only one semantic unit**... The other property is that morphs are simply stuck together ("agglutinated"); ideally, the **boundaries between morphs are clear**, and there is **no allomorphy**."

e. Whaley (1997: 134): "A language is *fusional* **if the boundaries between its morphemes are hard to determine**. The effect is as if the morphemes were blending, or fusing, together."

Most modern authors seem to agree that one of the ingredients of fusion is **cumulation**, i.e. the expression of several morphological subcategories in one affix (cf. the quotations from Bloomfield, Hjelmslev and Vance).[5] Agglutinating structures, by contrast, exhibit **separation** (to use a terminology that was introduced by Frans Plank, cf. Plank 1986, 1999).

Another characteristic of fusional languages is that they exhibit **stem alternations,** i.e. the (co-)expression of morphological categories by changing, rather than adding to, the stem (cf. Hjelmslev's definition). This is actually the oldest criterion: Friedrich von Schlegel and his 19th century followers were primarily struck by the difference between, on the one hand, languages like Sanskrit and German with their salient stem vowel changes, and on the other hand, languages like Quechua, which exhibit complete **stem invariance**.

A very vague criterion that is often mentioned in definitions of fusion and agglutination is that segmentation of morphemes is "difficult" (cf. Hjelmslev's, Anderson's and Whaley's definitions). The question of course is what properties of the language cause these difficulties. It seems that they often arise from stem alternations (see the last paragraph) or from **affix alternations**. The basic idea is that in a typical agglutinating language, each affix not only stands for just one subcategory, but is also invariant in its shape, whereas in fusional languages, not only stems, but also affixes show considerable morphophonological allomorphy.[6] Consider the partial paradigm of Hungarian noun inflection in (5).

---

[4] "the inflectional type, in which the boundary between root and suffix is not clear, in which each suffix often expresses several different grammatical properties simultaneously, and in which the root can undergo changes even in inflection."

[5] Note that I use the term *subcategory* for concepts such as singular, dative, future, and *category-system* for concepts like number, case, tense. (Plank 1999 uses the terms *term/category* in the same sense.)

[6] Greenberg (1954[1960:185]) uses morphophonological alternations of stems and affixes as the decisive criterion for his "index of agglutination", i.e. he disregards cumulation, and he lumps stem alternations and affix alternations together. (In this he is followed by other linguists working in the tradition of quantitative typology, such as Krupa 1965 and Silnitsky 1993.)

(5) Hungarian			‘house’	‘table’	‘r iver’
	Sg	Nom	*ház*	*asztal*	*folyó*
		Acc	*házat*	*asztalt*	*folyót*
	Pl	Nom	*házak*	*asztalok*	*folyók*
		Acc	*házakat*	*asztalokat*	*folyókat*

It is not immediately obvious how these forms should be segmented. For example, in the accusative singular form *házat*, the accusative suffix could be taken as *-at*, or as *-t* (as in *asztal-t* and *folyó-t*). In either case, we have to recognize an alternation, a stem alternation (*ház/háza*) or an affix alternation (*t/at*). Thus, it is the existance of alternations that makes segmentation difficult, and **affix invariance** makes it easy.

An even more radical kind of lack of one-to-one correspondence between meaning and form is **affix suppletion**, which is also commonly associated with fusional as opposed to agglutinating morphology (see especially Skalička 1951). Affix suppletion is allomorphy that cannot be described in phonological or morphophonological terms. It can be conditioned lexically, as in (6a); morphologically, as in (6b); or phonologically, as in (6c). Lack of affix suppletion will be called **affix uniformity**.

(6) a. Kannada (Dravidian; India)
Plural *-aru* /humans
-gaḷu /inanimates

b. Latin
1sg subject index *-o* /Present tense
-m /Imperfect tense

c. Lezgian (Nakh-Daghestanian; Caucasus; Haspelmath 1993:131)
Aorist Part. *-r* /after high vowel
-ji /after low vowel

It appears that **cumulation**, **stem alternation, affix alternation**, and **affix suppletion** are the key ingredients of non-agglutination or fusion. In the following, I will assume (somewhat counterfactually[7]) that no other morphological properties are relevant to defining the agglutination/fusion distinction.

---

[7] In a few works that address the agglutination/fusion distinction in some detail (Skalička 1951, Pöchtrager et al. 1998, Plank 1999, Plungian 2001), a number of further correlating properties have been mentioned:

agglutinating	**fusional**
affixes distinct	affixes often homonymous
always zero exponence	no/sporadic zero exponence
only local exponence	also extended exponence
repeatable affixes	unrepeatable affixes
large paradigm size	small paradigm size (cf. Plank 1986)
loose bonding	tight bonding
optionality	obligatoriness

Thus, the number of claimed correlations that can be found in the literature is actually much higher. However, since these correlations are little known, one would not say that using the terms “agglutination” and “fusion” implies accepting these correlations as valid. Thus, testing these further correlations is less urgent than testing the correlations that are implicit in the definitions of “agglutination” and “fusion”, and I will not say anything further about them here.

From the perspective of the individual properties, one could simply describe a language as cumulating, or as stem-alternating, or as affix-uniform, and so on. However, the fact that languages (or patterns) are typically described as "agglutinating" or as "fusional" shows that linguists expect the component properties to correlate with each other. So the Agglutination Hypothesis makes a second prediction:

(7) **Second Prediction: Correlation among different component properties**
If a language is agglutinating/fusional with respect to one of the component properties (cumulation, (stem/affix) alternation, affix suppletion), it will to behave similarly with respect to the other features.

Taken together, the two predictions made by the Agglutination Hypothesis are of course extremely interesting for the comparative linguist. But whether the terms "agglutination" and "fusion" are really useful depends also on whether the hypothesis is true. Somewhat surprisingly, this question has rarely been asked (Plank 1999, which came to my attention only after the first version of this paper had been finished, is the only exception known to me).

Apparently, it has not been generally recognized that the terms "agglutination" and "fusion" imply strong empirical claims, so linguists have not attempted to test these empirical claims. Another reason why such a test has not been carried out is probably that it encounters multiple difficulties, some of which will be addressed in the next section. However, I will also say how they can be overcome, so that in Section 4 I will present a first empirical test of the correlations.

3 Difficulties for an Empirical Test

If one wants to compare the morphological systems of widely different languages, one needs to make a large number of simplifications and idealizations, hoping that these do not introduce a bias into the investigation. In this section I discuss some of the difficulties I encountered in testing the two predictions of the preceding section, and make proposals for how to overcome them.

First, different languages show different morphological categories and vary significantly in morphological complexity. This is a difficulty especially for testing the First Prediction. For example, if this prediction is taken to imply that agglutinaton in nouns entails agglutination in verbs, it is impossible to test in languages that have inflectional morphology in verbs but not in nouns.[8] To address this difficulty, I only looked at languages with a fair amount of inflection, and took into account only the core of nominal and verbal inflection. Thus, I left aside inflection of adjectives, adpositions and other word classes (which are difficult to compare across languages

[8] Such languages could of course be taken to exhibit isolation in nouns, though this is rarely done.

anyway). I also left aside non-finite verb morphology, which does not belong to the core of verbal inflection and differs more across languages than tense, aspect and voice morphology. The assumption behind this is that if the idea at the heart of the Agglutination Hypothesis is correct, the correlations should emerge even if only the core inflectional categories of nouns and verbs are considered.

Second, it is well-known that cumulation is extremely common everywhere with person-number affixes (cf. Cysouw 2003: 296). Some examples of possessive person-number affixes in languages from different continents are given in (8):

(8)	Hungarian (Uralic; Europe)	O'odham (Uto-Aztecan; North America)	Tauya (Trans- New Guinea)	Lango (Nilotic; Africa)
Sg 1	*kez-em*	*ñ-kakkio*	*ya-neme*	*pàlà-ná*
2	*kez-ed*	*m-kakkio*	*na-neme*	*pàlə̀-ní*
3	*kez-e*	*kakkio-j*	*Ø-neme*	*pàlà-mɛ́rɛ̂*
Pl 1	*kez-ünk*	*t-kakkio*	*se-neme*	*pàlà-wá*
2	*kez-etek*	*'em-kakkio*	*te-neme*	*pàlə̀-wú*
3	*kez-ük*	*ha-kakkio*	*ne-neme*	*pàlà-gí*
	'my hand etc.'	'my legs etc.'	'my head etc.'	'my knife etc.'

Thus, including person-number affixes would lead us to find some amount of cumulation in the great majority of languages, even in languages that otherwise would probably be regarded as perfect examples of the agglutinating type. Thus, in my empirical test, I counted "person-number(-gender)" as a single inflectional category-system, so that these cases do not count as cumulation.

Third, in tense-aspect-mood forms, cumulation also seems to be common, but it is extremely difficult to identify the categories that are cumulated and to distinguish cumulation from semantic complexity. Consider the examples in (9).

(9)	**past + perfective**	**past + imperfective**	
a. Modern Greek	*é-γrap-s-e*	*é-γraf-Ø-e*	's/he wrote'
b. Lezgian	*qaču-na*	*qaču-zwa-j*	'took'
c. Italian	*port-ai*	*port-avo*	'I carried'
d. Pipil	*chiw-ki*	*chiwa-ya*	'did'

In Modern Greek, one can readily isolate a perfective marker $-s$ and a past marker *é-*. Both forms are fully compositional. In Lezgian (Haspelmath 1993), it is possible to identify an imperfective marker *-zwa* and a past marker $-j$, but the past perfective *-na* is not compositional. In Italian and Pipil (a Uto-Aztecan language of El Salvador), we have no compositionality at all. We might say that both Italian and Pipil are fully cumulative, but it is not even fully clear that "past" and "(im)perfective" are separate inflectional categories in these languages. To avoid these complications, I regarded "tense-aspect-mood" as a single category-system, unless there were very strong reasons to separate them.

Fourth, a problem with measuring the relevance of stem alternations is that in most languages, only a subset of lexical items show stem alternations (e.g. in German, only strong verbs alternate, and these are a small minority), and generally most lexical items do not alternate. We would probably want to quantify the importance of stem alternation: A language that has a stem change in one lexical item should not be regarded as equally fusional (with regard to the criterion of stem alternation) as a language that has stem changes in more than a hundred items. Thus, for each category-system separately, I assigned scores depending on the (estimated) absolute number of lexemes that exhibit stem alternations:

score of 1	fewer than 10 lexemes with stem alternation
score of 4	10–50 lexemes
score of 7	more than 50 lexemes
score of 10	the majority of (or all) lexemes

Fifth, morphophonological affix alternations are very difficult to distinguish from purely phonologically conditioned alternations. For example, the German plural suffix *-n/-en* (as in *Straße-n* 'streets', *Frau-en* 'women' *Partikel-n* 'particles') could either be described as morphophonological affix alternations (making German fusional in this regard), or as the same affix that happens to have different realizations for purely phonological reasons. After considering this problem thoroughly, I decided to exclude the property "affix alternation" from the empirical test, because it did not seem feasible to make the distinction in a coherent way for a greater number of languages.[9]

4 The Empirical Test

4.1 Data

Thus, in my empirical test of the two predictions I studied three properties of morphological systems: cumulation, stem alternation, and affix suppletion. I examined a sample of thirty languages which is somewhat biased in the usual way, with more languages from Europe than would be justified by the genealogical diversity in Europe. Not all the languages are unrelated, but each is from a different genus (in the sense of Dryer 2005). The thirty sample languages are listed in Table 1.

For each of these thirty languages, I determined a Cumulation Index, an Alternation Index, and a Suppletion Index, separately for nominal and for verbal inflection. The value of these indices is always between 0 (maximal agglutination) and 100 (maximal fusion). The Cumulation Index is defined as the percentage of inflected forms that exhibit cumulation; the Alternation Index is defined as the average

[9] I also simplified the procedure by considering only up to fifteen subcategories per lexeme class, and only up to ten subcategories per category-system, on the assumption that more was not required to capture the type of the language.

Table 1 The 30 languages of the sample

Niger-Congo	Swahili	Dravidian	Kannada
Nilotic	Lango	Tibeto-Burman	
Kordofanian	Krongo	Bodic	Classical Tibetan
Afro-Asiatic		Newaric	Dolakha Newari
Semitic	Arabic	Trans-New Guinea	
Egyptian	Coptic	Adelbert R.	Tauya
Indo-European		Madang	Amele
Germanic	German	Oceanic	Ponapean
Italic	Latin	Pama-Nyungan	Martuthunira
Iranian	Ossetic	Yuman	Maricopa
Indic	Hindi-Urdu	Uto-Aztecan	
Uralic		Numic	Tümpisa Shoshone
Finnic	Finnish	Aztecan	Pipil
Ugric	Hungarian	Mayan	Tzutujil
Lezgic	Lezgian	Cariban	Hixkaryana
Turkic	Turkish	Paezan	Páez
Tungusic	Evenki	Quechua	Huallaga Quechua
Yukaghir	Kolyma Yukaghir		
Nivkh	Amur Nivkh		

alternation score (cf. Section 3), i.e. the sum of stem alternation scores for each category-system divided by the number of category-systems; and the Suppletion Index is defined as the average percentage of subcategories (per category-system) that exhibit affix suppletion.

4.2 Testing the Second Prediction: Correlation Among the Three Indices?

Let us begin by examining the Second Prediction: is there a correlation among the three indices? Do languages with a lot of stem changes also show a lot of cumulation, and do languages with a lot of cumulation also show a lot of affix suppletion? The values of the three indices for the sample languages are shown in Table 2, where the languages are listed in rank order.

Comparison between the three indices is made difficult by the fact that the numbers stand for very different things. The total range of the values is not very different: between 0 and 66 for Cumulation, between 0 and 75 for Alternation, and between 0 and 84 for Suppletion. Thus, no language shows 100% fusion with respect to any of the indices, and for each index, there is at least one language with 100% agglutination. However, it is easily seen that the Cumulation values tend to be much lower than the Alternation and Suppletion values. Cumulation is simply a rare phenomenon outside of the Indo-European family, if cumulation of person and number is disregarded and if tense and aspect are not counted separately (see the discussion in Section 3). Only the Indo-European languages Hindi-Urdu and Latin have figures over 20% of cumulative forms.

Table 2 Index values and language ranking for each index

Ranking by Cumulation index:		Ranking by Alternation index:		Ranking by Suppletion index:	
1. Krongo	0	1. Hindi/Urdu	0	1. Nivkh	0
Lango	0	Hixkaryana	0	2. Tauya	3
Lezgian	0	Martuthunira	0	3. Ponapean	4
Newari	0	Nivkh	0	4. Quechua	10
Pipil	0	Páez	0	5. Páez	12
Ponapean	0	Swahili	0	6. Lezgian	12
Shoshone	0	Tauya	0	7. Tibetan	14
Tibetan	0	Turkish	0	8. Coptic	15
Turkish	0	9. Evenki	2	Krongo	15
Tzutujil	0	10. Coptic	5	10. Pipil	16
11. Swahili	0.1	11. Latin	7	11. Finnish	18
12. Amele	0.2	12. Tibetan	9	12. Maricopa	19
13. Maricopa	0.4	13. Krongo	12	13. Hixkaryana	22
14. Tauya	0.5	14. Yukaghir	13	14. Hungarian	23
15. Coptic	1.8	15. Maricopa	14	Turkish	23
16. German	2	Quechua	14	16. Swahili	28
17. Quechua	2.5	17. Lango	17	17. Martuthunira	29
18. Yukaghir	4	18. Newari	18	18. Lango	37
19. Hixkaryana	4.5	19. Tzutujil	24	19. Shoshone	38
20. Evenki	5	20. Amele	25	20. Evenki	39
21. Ossetic	6	21. Ossetic	26	21. Yukaghir	40
22. Nivkh	7	22. Shoshone	27	22. Newari	41
23. Arabic	8	23. Finnish	30	23. Hindi/Urdu	50
24. Finnish	13	24. Hungarian	36	24. Kannada	51
25. Kannada	14	25. Arabic	42	Ossetic	51
26. Hungarian	18	26. Kannada	43	26. German	56
27. Martuthunira	18	27. Lezgian	46	27. Arabic	62
28. Páez	30	28. Pipil	50	28. Amele	69
29. Hindi/Urdu	50	29. German	52	29. Tzutujil	77
30. Latin	66	30. Ponapean	75	30. Latin	84

Now the crucial question is: Do the languages tend to exhibit similar values for each of the three indices?[10] This is difficult to see by just inspecting Table 2, so in order to make the indices more comparable, I compared the rank values on the three indices. Table 3 lists languages in the order of average rank. The first three columns give the three indices, but the most important triple of figures is the rank values in

[10] That the different components of agglutination do not always go together has occasionally been noted in the literature. Thus, Vance (1987: 176) notes for Japanese:

> "Japanese morphology certainly tends to be agglutinative. The two properties involved in agglutination, however, do not correlate very well. Portmanteau morphs are rare… Allomorphy, on the other hand, is not at all uncommon…"

the next three columns. These figures show, for example, that Tauya has rank 14 for Cumulation, rank 4 for Alternation, and rank 2 for Suppletion, and so on. By the criterion of average rank, Tauya turns out to be the "most agglutinating" language, while Kannada is the "least agglutinating" language, and Indo-European languages such as Latin and German are also close to the bottom. This accords well with our expectations, of course, but there are also some surprises, especially perhaps the low (and hence "relatively fusional") position of Hungarian and Finnish. These languages have traditionally been labelled "agglutinating" (in Table 3, all languages which have been called "agglutinating" in the literature are preceded by an asterisk).

Of course, the confirmation of our expectations for Turkish and Latin does not mean that the Agglutination Hypothesis has been confirmed. On the contrary, it could well be that the models of these two well-known languages have been so powerful that linguists have unconsciously tended to define *agglutinating* as

Table 3 Index values and rank numbers for each language

	Indices			Ranks			(Average Rank)
	Cum	Alt	Sup	Cum	Alt	Sup	
Tauya	0.5	0	3	**14**	**4**	**2**	6.7
*Turkish	0	0	23	**5.5**	**4**	**14**	7.8
Tibetan	0	9	14	**5.5**	**12**	**7**	8.2
Krongo	0	12	15	**5.5**	**13**	**8.5**	9.0
*Nivkh	7	0	0	**22**	**4**	**1**	9.0
*Swahili	0.1	0	28	**11**	**4**	**16**	10.3
Coptic	2	5	15	**15**	**10**	**8.5**	11.2
Hixkaryana	4.5	0	22	**19**	**4**	**13**	12.0
*Quechua	2.5	14	10	**17**	**15.5**	**4**	12.2
Páez	30	0	12	**28**	**4**	**5**	12.3
*Lezgian	0	46	13	**5.5**	**27**	**6**	12.8
Ponapean	0	75	4	**5.5**	**30**	**3**	12.8
Maricopa	0.4	14	19	**13**	**14**	**12**	13.0
Lango	0	17	37	**5.5**	**17**	**18**	13.5
Pipil	0	50	16	**5.5**	**28**	**10**	14.5
Newari	0	18	41	**5.5**	**18**	**22**	15.2
*Shoshone	0	26	38	**5.5**	**22**	**19**	15.5
*Martuthunira	18	0	28	**27**	**4**	**17**	16.0
*Evenki	5	2	39	**20**	**9**	**20**	16.3
*Yukaghir	4	13.5	40	**18**	**14**	**21**	17.7
Tzutujil	0	24	77	**5.5**	**19**	**29**	17.8
Hindi-Urdu	50	0	50	**29**	**4**	**23**	18.7
*Finnish	13	30	18	**24**	**23**	**11**	19.3
Amele	0.2	25	69	**12**	**20**	**28**	20.0
*Hungarian	18	36	23	**26**	**24**	**14.5**	21.5
Ossetic	6	26	51	**21**	**21**	**24.5**	22.2
Latin	66	7	84	**30**	**11**	**30**	23.7
German	2	52	56	**16**	**29**	**26**	23.7
Arabic	8	42	62	**23**	**25**	**27**	25.0
*Kannada	14	42	51	**25**	**26**	**24.5**	25.2

* = language has been referred to as "agglutinating" in the literature.

"Turkish-like" and *fusional* as "Latin-like".[11] In this way, various properties that happen to be combined in these languages would have become part of the typological prototype, although there is in fact no tendency for these features to cooccur in languages cross-linguistically.

As can be seen from Table 3, there are some "well-behaved" languages that show similar rankings with respect to the three indices, e.g. Maricopa (ranks 13, 14, 12), Ossetic (ranks 21, 21, 24.5), Arabic (ranks 23, 25, 27), and Kannada (ranks 25, 26, 24.5). If all languages were like these, we would say that the three indices really cluster with each other, and that the Second Prediction of the Agglutinaton Hypothesis has been confirmed.

But unfortunately, too many languages are not "well-behaved". For example, Nivkh has the ranks 22, 4, 1, Páez has the ranks 28, 4, 5, Ponapean has the ranks 5.5, 30, 3, and Hindi-Urdu has the ranks 29, 4, 23. In fact, there is no statistically significant correlation among the three rankings: the Kendall coefficient of concordance is $W = 0.37$, $X^2 = 32.19$, and the significance is $p < .35$, well above the significance level of .05. My empirical test thus provides no support for the Second Prediction of the Agglutination Hypothesis.

This does not mean, of course, that I have definitively disconfirmed this prediction. My assignment of the various indices has involved some simplifications and arbitrary choices that could have influenced the results in a negative way. But my results present a strong challenge to those linguists who still want to maintain the agglutination-fusion distinction.

4.3 Testing the First Prediction: Correlation Among Nominal and Verbal Inflection?

But what about the First Prediction, concerning the correlation between different parts of the morphology? Since it is well-known that verbs and nouns sometimes behave differently (e.g. the Romance and Balkan Slavic languages lost their nomonal case inflection, but retained the synthetic verb inflection of their ancestors), one might not have particularly high expectations with regard to this prediction. Already Greenberg (1954[1960: 182]) noted that "a term like *agglutinative* applies primarily to a single construction. A language may well and indeed usually does contain some agglutinational as well as some nonagglutinational constructions." And Wolfgang U. Dressler, a consistent proponent of the agglutination-inflection typology along Skalicka's lines (cf. Dressler 1985, Pöchtrager et al. 1998), has recently acknowledged that "noun inflection and verb inflection may have a different typological character within the same language and develop diachronically in typologically different directions" (Dressler 2005: 7). Thus, it is not exactly expected that the First Prediction would fare better in my empirical test.

---

[11] A quotation such as the following is quite typical: "The ideal type of an agglutinating language, best represented in Turkish. . ." (Pöchtrager et al. 1998: 57).

To test it, I looked at nouns and verbs separately, for all three indices. Table 4 shows the index values and rank numbers, separately for the three indices.

We see that for Alternation and especially Suppletion, the rankings do correlate significantly (Spearman rank correlation coefficient, $p < 0.01$ (Suppletion), $p < 0.05$ (Alternation)). However, for Cumulation there is no significant correlation ($p < 0.20$).

Thus, we can say that languages tend to show (or lack) affix suppletion and stem alternations in both nouns and verbs simultaneously. In other words, languages can be meaningfully typologized as generally affix-suppleting (vs. affix-invariant) languages, and as generally stem-alternating (vs. stem-uniform) languages, at least across nouns and verbs. Thus, it is not completely meaningless to search for the kinds of correlations that we have been looking at here. But these positive correlations are hardly sufficient to justify a global distinction between "agglutinating" and

Table 4 Index values and rank numbers, separately for nouns (N) and verbs (V)

	Cumulation				Alternation				Suppletion			
	index		rank		index		rank		index		rank	
	N	V	N	V	N	V	N	V	N	V	N	V
Amele	0	0.5	**11**	**14**	0	50	**7.5**	**28**	91	47	**28**	**22.5**
Arabic	15	0	**25**	**6.5**	33	50	**22**	**28**	44	80	**24**	**29**
Coptic	0.6	3	**22**	**17**	10	0	**17**	**7**	6	24	**7**	**13**
Evenki	0	10	**11**	**23**	3	0	**15**	**7**	19	59	**14**	**25**
Finnish	19	7	**26**	**20**	42	18	**24**	**17.5**	21	14	**17**	**9**
German	0	4	**11**	**18**	35	70	**23**	**30**	67	45	**26.5**	**21**
Hindi/Urdu	100	0	**27.5**	**6.5**	0	0	**7.5**	**7**	100	0	**29.5**	**2**
Hixkaryana	0	9	**11**	**22**	0	0	**7.5**	**7**	20	24	**15.5**	**13**
Hungarian	1	36	**23**	**28**	60	13	**28**	**15**	13	33	**11**	**18.5**
Kannada	0	27	**11**	**26**	50	35	**26.5**	**23.5**	25	77	**19**	**28**
Krongo	0	0	**11**	**6.5**	0	25	**7.5**	**19**	20	10	**15.5**	**7.5**
Lango	0	0	**11**	**6.5**	0	33	**7.5**	**21**	27	47	**20.5**	**22.5**
Latin	100	33	**27.5**	**27**	0	14	**7.5**	**16**	100	69	**29.5**	**26**
Lezgian	0	0	**11**	**6.5**	50	42	**26.5**	**25**	8	17	**10**	**10**
Maricopa	0	0.7	**11**	**15**	17	12	**18**	**14**	6	32	**7**	**17**
Martuthunira	0	37	**11**	**29**	0	0	**7.5**	**7**	24	33	**18**	**18.5**
Newari	0	0	**11**	**6.5**	0	35	**7.5**	**23.5**	0	82	**3**	**30**
Nivkh	0	14	**11**	**25**	0	0	**7.5**	**7**	0	0	**3**	**2**
Ossetic	0	11	**11**	**24**	20	33	**19**	**21**	27	75	**20.5**	**27**
Páez	4	57	**24**	**30**	0	0	**7.5**	**7**	17	6	**13**	**4.5**
Pipil	0	0	**11**	**6.5**	66	33	**29**	**21**	7	24	**9**	**13**
Ponapean	–	0	**–**	**6.5**	100	50	**30**	**28**	0	7	**3**	**6**
Quechua	0	5	**11**	**19**	29	0	**21**	**7**	0	20	**3**	**11**
Shoshone	0	0	**11**	**6.5**	5	47	**16**	**26**	67	10	**26.5**	**7.5**
Swahili	0	0.1	**11**	**13**	0	0	**7.5**	**7**	50	6	**25**	**4.5**
Tauya	0	1	**11**	**16**	0	0	**7.5**	**7**	6	0	**7**	**2**
Tibetan	–	0	**–**	**6.5**	0	18	**7.5**	**17.5**	0	27	**3**	**15**
Turkish	0	0	**11**	**6.5**	0	0	**7.5**	**7**	15	31	**12**	**16**
Tzutujil	0	0	**11**	**6.5**	47	0	**25**	**7**	33	44	**23**	**20**
Yukaghir	0	8	**11**	**21**	27	0	**20**	**7**	31	50	**22**	**24**

"fusional" languages, because as we saw in the preceding section, affix suppletion and stem alternation do not correlate among each other.

5 Conclusions

The overall results of this study are mostly negative. The main positive point is that it is indeed possible, if difficult, to test the Agglutination Hypothesis empirically.

But the result of the preliminary empirical test carried out here is mostly negative. I have not found a statistically significant correlation between the Cumulation index, the Alternation index and the Suppletion index, which would have been expected if the correlations implicit in the composite types "agglutinating" and "fusional" existed.

Also, the evidence that nouns and verbs tend to behave alike is not overwhelming. There is no correlation with respect to cumulation, which is sometimes taken to be the primary defining property of "fusional" languages (especially in the more recent literature), and also for alternation, the evidence for a correlation is not particularly strong. However, with respect to affix suppletion, we can say that nouns and verbs tend to behave in the same way across languages.

Since the study is based on a relatively small number of languages and the sample is not truly representative of the diversity of the world's languages, these results are far from conclusive. I cannot say that I have shown that the Agglutination Hypothesis is wrong. However, the results cast sufficient doubt on the hypothesis to say that from now on, the burden of proof is on those who believe that it is correct.

As there are indications that stem alternation and affix suppletion characterize both nouns and verbs, one can legitimately characterize entire languages as *affix-suppleting* or *stem-alternating*. However, since cumulation does not tend to be similar across word classes, it is less meaningful to characterize entire languages as "cumulating" or "separatist". These terms are meaningful primarily when applied to particular morphological subsystems.

After this study, the terms *agglutination* and *fusion* have lost much of their legitimacy (unless they are given a technical sense that is at variance with the common usage). Of course, it is logically possible to define a composite type consisting of properties that do not correlate with each other (or only very partially), but it does not make much sense. While Turkish could perhaps still be characterized as "agglutinating" and Latin as "fusional", for many languages neither of these terms would apply, and it would not be possible to say that they are "intermediate" between these two extremes either. It is quite possible that the reason for the success of the agglutination/fusion distinction is that Latin and Turkish have been such prominent languages in Western linguistics over the last few hundred years. They differ strikingly in their morphological systems (cf. Plank 1991), and it is perhaps natural that from this point of view one would classify languages as more Latin-like or more Turkish-like. However, linguistics should move beyond Lationocentrism and Turkocentrism and try to do justice to each language, to describe and characterize it in its own terms, or in truly universal terms.

References

Anderson, Stephen R. 1985. Typological distinctions in word formation. In *Language typology and syntactic description, vol. III,* ed. Timothy Shopen, 3–56. Cambridge: Cambridge University Press.

Baker, Mark. 1996. *The polysynthesis parameter*. New York: Oxford University Press.

Bauer, Laurie. 1988. *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.

Bazell, Charles. 1958. Linguistic typology. In *Five inaugural lectures*, ed. Peter D. Strevens, 29–49. London: Oxford University Press.

Bickel, Balthasar and Johanna Nichols. 2005. Inflectional synthesis of the verb. In *The World Atlas of Language Structures*, ed. Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard, 94–97. Oxford: Oxford University Press.

Bloomfield, Leonard. 1933. *Language*. New York: Holt & Co.

Cysouw, Michael. 2003. *The paradigmatic structure of person marking*. Oxford: Oxford University Press.

Dressler, Wolfgang U. 1985. Typological aspects of Natural Morphology. *Acta Linguistica Hungarica* 35: 51–70.

Dressler, Wolfgang U. 2005. Morphological typology and first language acquisition: Some mutual challenges. In *Morphology and Linguistic Typology: On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania, 21–23 September 2003*, ed. Geert Booij, Emiliano Guevara, Angela Ralli, Salvatore Sgroi and Sergio Scalise, 7–20. (http://mmm.lingue.unibo.it)

Dryer, Matthew S. 2005. Genealogical language list. In *The World Atlas of Language Structures*, ed. Haspelmath, Martin & Dryer, Matthew S. & Gil, David & Comrie, Bernard, 584–644. Oxford: Oxford University Press.

Evans, Nicholas. 1995. *A grammar of Kayardild*. Berlin: Mouton de Gruyter.

Greenberg, Joseph. 1954 [1960]. A quantitative approach to the morphological typology of language. In *Method and Perspective in Anthropology: Papers in Honor of Wilson D. Wallis*, ed. Robert F. Spencer, 192–220. Minneapolis: University of Minnesota Press. [Reprinted in: *International Journal of American Linguistics* 26 (1960): 178–94.]

Greenberg, Joseph. 1974. *Language typology: A historical and analytic overview*. The Hague: Mouton.

Haspelmath, Martin. 1993. *A grammar of Lezgian*. Berlin: Mouton de Gruyter.

Hjelmslev, Louis. 1963. *Sproget: En introduktion*. København: Berlingske Forlag. (German translation: *Die Sprache*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1968)

Humboldt, Wilhelm von. 1822/1985. *Ueber das Entstehen der grammatischen Formen, und ihren Einfluss auf die Ideenentwicklung*, reprinted in: *Über die Sprache: Ausgewählte Schriften*, ed. Jürgen Trabant, München: Deutscher Taschenbuch Verlag.

Humboldt, Wilhelm von. 1836. *Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluß auf die geistige Entwickelung des Menschengeschlechtes*. Reprinted.

Hyman, Larry. 2001. Tonal systems. In *Language typology and language universals: An international handbook* (HSK), vol. II, ed. Martin Haspelmath et al. Berlin: de Gruyter, 1367–1380.

Krupa, Viktor. 1965. On quantification in typology. *Linguistics* 12: 31–36.

Müller, Friedrich Max. 1871. *Lectures on the science of language*. 2nd. Ed. New York: Scribner.

Plank, Frans. 1986. Paradigm size, morphological typology and universal economy. *Folia Linguistica* 20: 29–48.

Plank, Frans. 1991. Of abundance and scantiness in inflection: a typological prelude. In *Paradigms: the economy of inflection*, ed. Frans Plank, 1–39. Berlin: Mouton de Gruyter.

Plank, Frans. 1998. The co-variation of phonology with morphology and syntax. *Linguistic Typology* 2.2:195–230.

Plank, Frans. 1999. Split morphology: How agglutination and flexion mix. *Linguistic Typology* 3.3: 279–340.

Plungian, Vladimir A. 2001. Agglutination vs. flection. In *Language typology and language universals: An international handbook* (HSK), vol. I, ed. Martin Haspelmath et al., 669–678. Berlin: de Gruyter.

Pöchtrager, Markus A. & Bodó, Csanád & Dressler, Wolfgang U. & Schweiger, Teresa. 1998. One some inflectional properties of the agglutinating type illustrated from Finnish, Hungarian and Turkish inflection. *Wiener Linguistische Gazette* 62-63: 57–92.

Sapir, Edward. 1921. *Language*. New York: Harcourt.

Schlegel, August Wilhelm von. 1818. *Observations sur la langue et la littérature provençales*. Paris: Librairie Grecque-latine-allemande.

Schlegel, Friedrich von. 1808. *Ueber die Sprache und Weisheit der Indier: Ein Beitrag zur Begründung der Alterthumskunde*. Heidelberg: Mohr & Zimmer.

Schleicher, August. 1850. *Die Sprachen Europas in systematischer Übersicht*. Bonn: H. B. König.

Silnitsky, George. 1993. Typological indices and language classes: a quantitative study. *Glottometrika* 14: 139–160.

Skalička, Vladimír. 1951. Das Erscheinungsbild der Sprachtypen. In Skalička, Vladimír. 1979. *Typologische Studien*. Braunschweig: Vieweg, 21–58.

Slobin, Dan. 1997. The origins of grammaticizable notions. In *The Crosslinguistic Study of Language Acquisition, vol. 5: Expanding the Contexts*, ed. Dan Slobin, 265–323. Mahwah, NJ: Erlbaum.

Spencer, Andrew. 1991. *Morphological theory*. Oxford: Blackwell.

Testelets, Yakov G. 2001. Russian works on linguistic typology in the 1960s–1990s. In *Language typology and language universals: An international handbook*, vol. I, ed. Martin Haspelmath et al., 306–323. Berlin: de Gruyter.

Vance, Timothy J. 1987. *An introduction to Japanese phonology*. Albany: SUNY Press.

Whaley, Lindsay J. 1997. *Introduction to typology*. Thousand Oaks/CA: Sage.

What Linguistic Universals Can be True of

Aditi Lahiri and Frans Plank

Abstract Universals in linguistics were traditionally intended to be true of languages: "for all languages, p" or "for all languages, if p then q". Our contention, by contrast, is that many universals have a narrower scope than languages as such, or mental lexicons-and-grammars as such. Linguistic universals are not axiomatically to be conceived of as universals of language: it is only derivatively—namely if universals are true of all parts of each language and of all representations of forms-in-constructions of each language—that this is what they may amount to. Only very basic organising principles of lexicons and grammars should really be expected to make their influence felt pervasively, over all parts and all representations.

Keywords Adjective order · asymmetry · coronal · gender · infixation · markedness · universals

1 Introduction

What universals in linguistics were traditionally intended to be true of was languages: "for all languages, p" or "for no language, not p", or, in the case of covariation, "for all languages, if p then q" (or, equivalently "if not q then not p" or, restated non-implicationally, "not (p and not q)") or "for no language, if p then not q" (or "if q then not p" or "not (p and not q)"). However time-honoured this manner of speaking and thinking, for many p's and q's it suggests too global a scope for constraints on linguistic diversity. It can be deceptive on several grounds to axiomatically equate "linguistic universals" with "universals of language".

A. Lahiri
University of Oxford

F. Plank
Universität Konstanz

S. Scalise et al. (eds.), *Universals of Language Today*, 31–58,
DOI: 10.1007/978-1-4020-8825-4_3,

2 True of All Languages

It is not only difficult but impossible to prove universal claims—linguistic universals as well as cognitive or phonetic universals in which they may be grounded—which are meant to be true of the domain of ALL human languages EVER; empirically true, that is, rather than true by definition of "language".[1] And when they are statistical rather than absolute it is equally impossible to DISprove universal claims of this intended scope. For many languages which are known to exist or to have existed knowledge is inaccessible, limited, conjectural, or forever irretrievable; and outside a narrow window in the history of speaking man—some 10,000 years at best out of so far 100 or 200,000—no knowledge about languages is to be had at all. Hypotheses about universals have to be formed, and their validity has to be ascertained, without any regard to languages outside this window, on the side of both the past and the future (assuming *homo loquens* has one). Therefore, the empirical basis of universals research can only be (a sample of) a subset of the domain for which universals could maximally claim validity, and have traditionally been claiming validity: that of humanly possible languages. For several reasons, there is no way of dependably generalising from (a sample of) this subset to the entire domain, however uniformitarian one's convictions: (i) linguistic diversity is likely to have been massively reduced in the distant past through natural catastrophes decimating mankind, and unlike with life forms there are no fossil traces allowing us to infer what happens to have been lost; (ii) patterns of linguistic diversity have been randomly skewed throughout human history insofar as they were concomitant to the histories of populations—migrating or stagnant, conquering or conquered, surviving or extinguished, culturally influential or influenced; (iii) at any point in time, future linguistic diversity has been, and will be, a function of linguistic diversity of the past insofar as learners, however linguistically imaginative, cannot but model their own languages on only those languages that are still around to be learned.

To seriously confirm or disconfirm universals in the strict and most ambitious sense, it would need several reruns of the linguistic history of *homo sapiens sapiens* under randomly differing population-historical conditions, so as to be able to see how diversity and unity would be recreated through innovations and retentions in 4,000–8,000 successive generations of learners altering or preserving the human protolanguage (itself still the chimera which it was in 1866 when the Société Linguistique de Paris wisely ruled the origin of language unfit as a topic for scholarly publication).

---

[1] For those who require universals only to be POTENTIALLY true of languages, rather than ACTUALLY true, it is easy to establish universality: anything attested in one individual language (clicks, labial affricates, stem-internal reduplication, verb-second, a basic colour term for 'turquoise', etc., whatever deserves exhibition in DAS GRAMMATISCHE RARITÄTENKABINETT at http://typo.uni-konstanz.de/rara/intro/) is universal in this sense. The corresponding problem here is that there is no way of even guessing whether you know 1% or 99% of the contents of this common fund from which particular languages are making their choices.

The only viable domain for universals research, then, is all-languages-present-and-past-as-known-to-us-now. The risk, inevitably, is that this subset is atypical of the set of all-languages-ever: what appear to be "universals" may, *sub specie aeternitatis*, only be historical contingencies.

3 True of All Varieties

It may seem immodest but is really trivial: we would in fact want universals not only to be true of languages-as-known-to-us-now, but of each variety—dialect, sociolect, idiolect, register—of each known language. Naturally, if one variety of some language had some property *p* which another variety of the same language, and perhaps the variety designated as a standard and codified in reference grammars, was known to be lacking, no one would claim universal status for *p*. There are no, say, dialect universals, intended as valid only for dialects, as distinct from standard-language universals. (Which is not to say there can be no universals concerning the range of possible divergences from a norm; but that would be a diachronic issue.)

Fortunately, distinguishing what is a language and what is a variety of a language isn't one of your problems when you are in the universals business. Assuming there are lexicons-and-grammars mentally represented by indivduals who on this basis perform speech acts which co-members of their speech communities can make sense of through their own mental lexicons-and-grammars, the individuals whose diversity and unity across mankind-as-known-to-us we are in the business of studying are individual mental lexicons-and-grammars—ALL of them about which knowledge is to be had. Of course, if you had a hard time trying to representatively sample languages, and are now being asked to sample mental lexicons-and-grammars instead, your practical problems will be enormous; but this is a different matter.

What matters for present purposes is the recognition that it is individual lexical and grammatical innovations which bring about linguistic diversity—those, that is, which prove socially successful, diffusing through speech communities or segments of them, effectuating change. And wherever diversity is limited through universals, it must be individual innovations, or sets of them in the case of co-variation, which are subject to constraints. It is not entire languages that are being innovated at one go.

4 True of All Speech Acts

Lexicons-and-grammars manifest themselves in speech acts: disregarding extraneous interferences, no single speech act, in any known variety of any known language, should therefore be violating any valid universal constraint on lexicons-and-grammars. Although this expectation is trivial, too, the proviso needs to be added that only such properties of speech acts are at issue as are subject to the regulation of lexicons-and-grammars.

For example, segment inventories universally distinguish between consonants and vowels, and no variety of any known (spoken) language is at odds with this universal, interpreted phonetically or phonologically.[2] However, not all utterances to be heard in all languages—of words or whole sentences—will contain both consonants and vowels, phonetically or also phonologically: *xɬp'χʷɬtɬpɬɬs kʷc'* 'then he had had in his possession a bunchberry plant' or *c'ktskʷc'* 'he arrived' from Bella Coola only consist of consonants (some syllabic); *uouoa* 'false mullet' (a fish) or *aːeːiː eia* (ending of chant) from Hawaiian only consist of vowels (Hyman 2008). Given the segment inventories of these two languages, one Salishan and the other Austronesian, and the way the available segments are deployed in lexical entries, it is not entirely accidental that vowel-less or consonant-less utterances should occasionally occur here (rather than, say, in English). Nonetheless, such occurrences are not as such rule-governed (that is, vowellessness or consonantlessness does not figure in any rule or constraint of the grammars of the languages concerned), and thus are not under the direct jurisdiction of constraints on linguistic diversity and on linguistic innovations which bring it about.

5 The Scope of Universals, Often Limited

With diversity ultimately due to individual innovations, the questions to be addressed, then, are these: Is it really WHOLE languages (as we know them), or WHOLE mental lexicons-and-grammars (as we are able to describe them), that universals are to be expected to be true of? And what if they do happen to be true of WHOLE languages, or WHOLE mental lexicons-and-grammars, but are not also true of some of their PARTS?[3] In reply, our contention is that languages as such, or mental lexicons-and-grammars as such, are not the right kind of individuals to generalise over in universals research, or at any rate to begin to generalise. LINGUISTIC universals are not axiomatically to be conceived of as universals OF LANGUAGE: it is only derivatively—namely if universals are true of ALL PARTS of each (known) language and of ALL REPRESENTATIONS of forms-in-constructions of each (known) language—that this is what they may amount to. Individual innovations, bringing about diversity, may well be rather local in scope and affect some parts and some representations independently of others, and may be independently reined in by constraints of their own. Upon reflection, it is only very basic organising principles of lexicons and grammars which should really be expected to make their influence felt pervasively, over all parts and all representations.

---

[2] For an extensive and ongoing documentation of universals, including those discussed and not specially referenced in this paper, see THE UNIVERSALS ARCHIVE at http://typo.uni-konstanz.de/archive/intro/.

[3] Or, as one reviewer puts it, a bit more formally: What if a constraint only holds of a proper subset, but not of a set as a whole? The point of Section 4 above was that individual speech acts are not such subsets, of grammatical relevance, that this sort of question would be meaningfully asked about them.

6 True of only Some Language-Parts

First, it may be the case that universals-intended-as-valid-for-known-languages are valid or invalid depending on whether they are intended to apply to LANGUAGES AS SUCH or to all PARTS of them—that is, all particular words in all their forms in all their particular constructions.

6.1 Some Words as Misfits

To see that this can make a difference, take the universal, often claimed to be unexceptional, that gender distinctions are unequally distributed over numbers, favouring the singular over non-singular numbers. This is an instance of the more general universal, widely invoked, that marked terms of morphological categories (such as plural or dual numbers) are more disposed than their unmarked opposites (such as singular) to license neutralisations of term-distinctions for categories they intersect with.[4]

If its lexicon and grammar, especially its inflectional morphology, are considered in their entirety, Spanish conforms to the universal, stated implicationally, that if a gender distinction is found in non-singular numbers, it will also be found in the singular: Spanish distinguishes two genders (masculine and feminine) in both singular and plural, and a marginal third ("neuter") only in the singular. However, there are two words in Spanish, the independent 1st and informal 2nd person personal pronouns, which distinguish masculine and feminine only in the plural—consisting of original 1st/2nd person plural pronouns *nos/vos*, to which the adjective *otr-os/-as* 'other' has come to be added, which retains the gender contrast of adjectives—but not in the singular:

(1) Spanish personal pronouns (only subject forms given)

		SG			PL	
		MASC	FEM	NEUT	MASC	FEM
1st		------- *yo* ------			*nosotros*	*nosotras*
2nd	INFORMAL	------- *tú* -------			*vosotros*	*vosotras*
	FORMAL	----- *usted* -----			-------- *ustedes* --------	
3rd		*él*	*ella*	*ello*	*ellos*	*ellas*

Similar patterns are not uncommon elsewhere. Sometimes gender distinctions are found to be inoffensively distributed over numbers when the languages concerned are looked at as a whole; but then some words in these languages, namely pronouns, preferably of 1st and 2nd person, as in Lithuanian (2), are seen to limit gender distinctions to the dual (which typically includes the numeral 'two', continuing to

[4] For extensive documentation see Plank & Schellinger 1997, and, as always, THE UNIVERSALS ARCHIVE.

agree in gender in the manner of a lower numeral after being grammaticalised as part of a dual form), while the corresponding words in the singular—'I' and 'you'—lack this gender contrast.[5]

(2) Lithuanian personal pronouns (only nominative forms given)

	SG		DU		PL	
	MASC	FEM	MASC	FEM	MASC	FEM
1st	------ *àš* ------		*mùdu*	*mùdvi*	----- *mẽs* ----	
2nd	------ *tù* ------		*jùdu*	*jùdvi*	----- *jũs* -----	
3rd	*jìs*	*jì*	*juõdu*	*jiẽdvi*	*jiẽ*	*jõs*

Another familiar morphological universal is familiarly intended as valid for whole languages: If a dual (a highly marked number) is distinguished in number systems, then a plural (less marked) will also be distinguished. While the universal is probably valid for all known languages, notwithstanding some where a plural would seem somewhat less entrenched than the dual in terms of frequency of occurrence or indeed of obligatoriness, it is not so uncommon for some words in languages with a dual, namely words designating natural pairs, to only inflect for dual but not for plural (Plank 1989: 317–318). Thereby, a semantically coherent subset of words violates a universal that remains true of the respective languages as a whole.

As a phonological example, consider a familiar instance of asymmetry in phoneme inventories (and see further Section 8 below): universally, "umlauted" vowels (front, round /y, ʏ, ø, œ/) imply their unumlauted counterparts (/u, ʊ, o, ɔ/). New High German has both series (with perhaps a tendency in dialects to unround umlaut vowels), and as a language therefore does not give offence; neither did earlier Germanic prior to umlauting, showing only the unumlauted series of vowels. When umlauting began, the corresponding vowel series used to be distributed unequally over the inventory of form classes: only stressed syllables of content or function words could have their vowel umlauted by a following high vowel or glide; elsewhere—that is, in affixes, whether stressed or unstressed, as well as in unstressed syllables of words—only unumlauted vowels would be found. Originally, therefore, as long as umlaut was a productive phonological process, each of these form classes individually would conform to the universal, too: there were either only unumlauted vowels, or unumlauted as well as umlauted vowels; no form class had only umlauted vowels. Then, with the phonological umlaut triggers gone and with umlaut partly morphologised, the distributions over form classes changed: now there are words

[5] In other, ostensibly similar instances of particular words offending against this universal there are no other words which would exculpate the languages concerned. Sometimes words which are used to refer to sets of referents of potentially mixed gender, namely anaphoric pronouns ('they' as 'he and she' as opposed to 'he and he' and 'she and she'), may have an extra gender form in their plural vis-à-vis the corresponding singular (rather than practising gender resolution in favour of the unmarked gender, as in French, or also rather than neutralising gender entirely in non-singular, as in English); naturally, no words will have SINGULARS of mixed gender, a contradiction in terms.

(both content and function words, such as *Tür* 'door', a noun, *schön* 'beautiful', an adjective, *stöhnen* 'groan', a verb, and *für* 'for', a preposition) which have umlauted vowels lacking an unumlauted counterpart — which is precisely the asymmetry proscribed by the implication at issue. It is (i) words where umlaut has not been generalised and unumlauted and umlauted vowels alternate across inflectional paradigms and/or between base and derivation (morphologically conditioned) and (ii) words wholly without umlaut which bring the vowel inventory of New High German up to standards, as defined by the implication that umlauted imply unumlauted vowels.

6.2 Some Word-Forms as Misfits

In a variation on this theme of exceptional parts (words) of unexceptionable wholes (languages), when all individual words, or rather lexemes, of a language behave as dictated by a universal, it may still be the case individual inflectional FORMS of some words misbehave.

With a symmetric gender system of masculine, feminine, neuter distinguished in both singular and plural, and with more gender neutralisations in plural than in singular, Latin, as a language, conforms to the universal about permissible gender-number skewings. However, there are several sets of words in Latin which inflect for gender, number, and case and distinguish genders in both singular and plural, but which deviate insofar as only in certain cases genders are distinguished in the plural but not in the corresponding singular case form. This more circumscribed kind of deviation is found, for example, in the nominative with all present participles and the so-called adjectives of one termination (such as *fēlīk-s* 'happy'), where neuter is distinct from masculine/feminine only in the plural (3); in the genitive with all words following the pronominal inflection (such as the proximal demonstrative *is*, *ea*, *id* 'he, she, it; this'), where masculine/neuter is distinct from feminine only in the plural (4); in the accusative with *o-/a*-stem adjectives (such as *māgn-us* 'great'), where all three genders are only distinct in the plural (5).

(3) Latin adjectives of one termination (partial)

	SG			PL		
	MASC	FEM	NEUT	MASC	FEM	NEUT
NOM	---------- *fēlīk-s* ----------			----- *fēlīk-ēs* -----		*fēlīk-ia*
ACC	---- *fēlīk-em* ---	*fēlīk-s*		----- *fēlīk-ēs* -----		*fēlīk-ia*

(4) Latin pronominal inflection (partial)

	SG			PL		
	MASC	NEUT	FEM	MASC	NEUT	FEM
NOM	*is*	*id*	*ea*	*iī*	*ea*	*eae*
GEN	---------- *ē-ius* ----------			----- *ē-ōrum* -----		*ē-ārum*

(5) Latin *o*-/*a*-stem adjectives (partial)

	SG			PL		
	MASC	NEUT	FEM	MASC	NEUT	FEM
NOM	*māgn-us*	*māgn-um*	*māgn-a*	*māgn-ī*	*māgn-a*	*māgn-ae*
ACC	------ *māgn-um* -------		*māgn-am*	*māgn-ōs*	*māgn-a*	*māgn-ās*

As a possible phonological analogue, assume that if the cluster [nz] is permitted word-finally, then [ns] will also be permitted in this position, provided there is a voicing contrast for sibilants to begin with. Due to the frequency of word-final devoicing in the world's languages, this would seem not too wildly implausible a universal to assume. Modern English as a whole, a language not given to final devoicing, would be inoffensive, with both [nz] and [ns] found word-finally. All words of English, in their basic form, behave as demanded by the putative universal, too, which only excludes the constellation "permissible [nz], impermissible [ns]": with no (synchronic) morpheme boundary between the segments of the cluster, only [ns] occurs, but not [nz] (e.g., *dense*, *tense*, *sense*, *fence*, *flense*, *manse*, *dance*, *lance*, *trance*, *rinse*, *prince*, *wince*, *since*, *science*, *disturbance*, *conference*, *Hanse*, *trans-*).[6] However, when a morpheme boundary intervenes—a transparent one, that is: hence [ns] equally in *cleanse*, *pence*, *(w)hence*, *(n)once*, *summons*, or the exclamation *good heavens!*, non-transparently related to *clean*, *penny*, *(w)hen*, *(n)one*, *summon*, *heaven*—[nz] is the only surface option, with voicing assimilation obligatory in derived environments (plural *pen-s*, *heaven-s*, 3rd singular *win-s*, genitive *the nun's tale*, clitic copula *the hen's here*). Inflected word-forms and host-enclitic constructions, thus, show the very constellation that is prohibited, "permissible [nz], impermissible [ns]".

6.3 So What?

When some parts of some (or perhaps also many) languages are found to be at odds with a universal that holds true for other parts, however close the affinity between the parts might seem with respect to the phenomenon concerned, it would be as rash to unceremoniously drop that universal or downgrade it from a categorical to a statistical universal as it would be complacent to accept it as confirmed, valid as it is for languages as a whole. Rather, the conclusion to be drawn is that universals are not perforce to be predicated of languages as a whole. And the question to be confronted is why some parts of lexicons-and-grammars are exempt from constraints limiting the diversity of other, kindred parts.

---

[6] Proper names like *Jones*, *Stevens*, *Collins*, *Orleans* permit [nz], with final *-s* perhaps analysed as a naming formative. So do *lens*, perhaps continuingly analysed as a plural of sorts (despite the agreement); a few loans or technical terms like *bonze*, *bronze*, *contredanse*, *winze*; anglified Latin expressions like *locum tenens* or *homo sapiens*; partly also prefix *trans-*.

As to the examples discussed in this section, gender distinction is not a property of languages: it is particular words or word classes which distinguish gender and trigger gender agreement and which, in particular inflectional forms, agree in gender; on present evidence, innovating, maintaining, altering, abandoning gender distinctions can to some extent be done independently from one word or word class to another, or from one inflectional category to the other. Universals constrain the extent of this independence; but they must not constrain it too much. No gender distinction must be forced on 1st and 2nd person singular pronouns, most reluctantly gendered anywhere, whenever dual or plural 1st/2nd person pronouns have inadvertently gained themselves one through grammaticalisation. No gender distinction must be forced on words and their singular inflectional forms in Indo-European languages such as Latin when the case-marking pattern and exponents specifically for neuters prevails in the plural, overriding all other paradigmatic design specifications. Things may happen to parts of inflectional paradigms for all sorts of reasons, morphological, phonological, or syntactic, and symmetric or suitably asymmetric gender distinction may not always have the highest priority in actuating or counteracting paradigmatic changes.

Dual marking is not a property of whole languages, either: it is particular words and word classes which, in particular inflectional forms, inflect for this number category, and perhaps agree in dual. Again, on present evidence, innovations, maintenance, alterations, and losses of duals in relation to other numbers can to some extent proceed independently from one word or word class to the other. However rigidly the extent of this independence is constrained, no plural must be forced on natural pair nouns whenever they have innovated a dual, the number most congenial to them.

Phonotactic constraints are not properties of whole languages, either: they may selectively apply or not apply to different kinds of domains, such as simple words and complex words. On present evidence, when progressive voicing assimilation only applies in derived environments of complex words, then contrasts in the final segment of clusters that would obtain otherwise, with voiceless preferred word-finally, must be allowed to be reversed.

Lastly, parts of segment inventories can be deployed selectively across the form classes of a language. With umlauted front, rounded vowels in Germanic this distribution over form classes and positions in words follows from the way this whole series was innovated. With umlaut susceptible to be morphologised, on present evidence, the umlaut series must be allowed to become independent of its unumlauted counterpart series and lead a distributional life of its own.

In sum, it is an empirical issue to determine what the minimal units and processes are that can vary independently of one another from one lexicon-and-grammar to the other: one domain of occurrence of segments and clusters from others in phonology; one word class from other word classes, one word of one word class from other words of the same class, one inflectional form from other inflectional forms of the same word in inflection; one construction from other constructions, one form-in-construction from other forms in the same construction, rules of construction from other rules in syntax. It is a further empirical issue to ascertain co-variation for

such elementary units and processes, and thereby to seek to generalise as far as possible, but not further. The challenge is to find reasons why not all individual parts always do as the strongest generalisations over whole languages would have it: sometimes, this reason is that non-conforming parts, newly grammaticalised or otherwise innovated, or left-overs of other parts recently lost, have not had enough time yet to come to conform.

7 True of only Some Representations of Forms-in-constructions

Given that lexical forms and grammatical constructions are complex structures with potentially several representations, these representations may be constrained selectively, each according to its own rationale. Refining traditional approaches where universals were assumed to be globally true of forms-in-constructions per se, universals have more recently been distinguished as "descriptive" and "analytic"/"architectural" (Hyman 2008), as "deep" and "surface" (Newmeyer 2008), "concrete" (Plank 2007), depending on which kinds of representations they specifically constrain.

7.1 Morphology Disarrayed when Pronounced

As a first example, consider infixes. Their relevance here is that morphological constituents, in particular stems, may be licensed to be discontinuous in representations of words close to the way they are pronounced, whereas in morphological representations themselves, close to the way complex meanings are constructed, such discontinuities are prohibited.[7]

On the face of it, affix order is evidently variable: there are suffixes, prefixes, circumfixes (perhaps including "parafixes", consisting of an adfix and an infix), transfixes (unlike circumfixes morphologically not bipartite), and infixes—the first by far most common and the last by far least common crosslinguistically. An implicational universal can be stated (and has been stated) which would account for the unequal crosslinguistic distributions, insofar as the availability of one kind of affixes is made conditional on the presence of other kinds: If there are infixes, there will also be adfixes (= suffixes and/or prefixes, also circumfixes other than parafixes). Achronically, this constraint can be seen as a particular case, pertaining to stems, of a more general constraint against discontinuous constituents: Discontinuous constituents are dispreferred as being more difficult to store, access, and process than continuous constituents. This constraint can also be seen as derivative of a diachronic law: Adfixes are the only productive source of infixes, and phonological rearrangement is the only productive mechanism of getting adfixes inside

[7] For epic surveys of infixation see Ultan 1975, Moravcsik 1977, 2000, Yu 2007. A short story has been distilled from them in Plank 2003, 2007, now slightly rewritten in light of Yu 2007.

stems (in the right phonological circumstances, with adfixes remaining external elsewhere). There are in fact one or two other scenarios for the emergence of infixes, in particular "entrapment" and "excrescence": when an outer adfix or function word is reanalysed as part of the stem, an inner affix may thus be trapped in between the two parts of new bipartite stems; an internal syllable or sequence of segments, originally meaningless but found to recur in several stems, may be reanalysed as a morphological word-part and eventually get inserted in other stems, too.[8] At least in the case of entrapment, infixes imply adfixes, too, diachronically and very likely also synchronically.

An even stricter constraint would be not to permit infixation at all, anywhere and at any time, rather than allowing it on the (achronic or diachronic) condition that there is also adfixation. Though it is desirable to push constraints as far as possible, this move would seem glaringly at odds with crosslinguistic reality: there ARE affixes inside stems, and they DID get inside stems from external origins. However, when word representations are separated into an abstract morphological one, taking care of the construction of complex meanings and forms from their component parts, and a concrete one that is to be pronounced, the strict constraint does prove tenable if selectively imposed on morphological representations. In thereby absolving the morphology of the responsibility for infixes, at least those originating through metathesis, light is also shed on constraints on infix constructions that would otherwise seem accidental. With the historical origin of affixes overwhelmingly external, owed to the univerbation of separate contiguous words (and only rarely to excrescence), morphology prefers order to be rigid: (i) reorderings of affixes among each other and relative to stems are rare, and may need special licensing by prosodic or scope-semantic considerations; (ii) partial interlacings of external addenda with stems are morphological anathema—unless the fault lies with stems themselves, as when bipartite stems are being created, trapping adfixes in between their parts, now infixes of sorts. It would certainly be odd if serial order in morphological representations were completely ignored when complex words are pronounced; nonetheless, the units structuring pronunciation are not identical to morphological ones, and for its own purposes pronunciation (or indeed also perception) may find divergent representations preferable, with relevant parts arranged differently. Arguably, then, infixations (at least if not due to entrapment) are the sole responsibility of phonology, which is not bound to maintain continuous morphological constituency should other considerations prevail; "infixes" (other than those entrapped) are adfixes attached to phonological rather than morphological units.

---

[8] Estimates differ on whether entrapment is exceedingly rare or not uncommon. The American English infix *-ma-* (as in *sophisti-ma-cated*) supposedly illustrates excrescence, being traced to /ma/'s in filler words such as *whatchamacallit* (< *what you may call it*) and *thing(a)mabob* (< *thingum(a)-bob*) by Yu (2007: 174–177). Yu (2007: 157–172) also suggests "reduplication mutation" — complex internal reanalyses of opaque reduplicative constructions — as a further scenario of infix emergence.

The relevant prosodic units and structures which are at a premium in phonological arrangements, including those of morphological representations with adfixes (with those affected underlined in examples), are these, and only these:

(i) SYLLABLES, as syllabified at the stem or perhaps also the word level, with the arrangement of segments aiming at CV patterns, at permissible clusters, at having clusters internally rather than at word-edges, at sonority sequencing, or at overall syllabic compactness.

Examples:

- perfective verbs in Tagalog (Malayo-Polynesian, Austronesian) such as *k-um-ain* 'ate', *p-um-asok* 'entered';
- present stems of verbs of the relevant conjugation class in Latin (Italic, Indo-European) such as *fu-N-d-* 'shed', *vi-N-k-*'conquer', *ru-N-p-*'break' (with the nasal assimilating in place to the following consonant);[9]
- nominalisations of verbs in Leti (Malayo-Polynesian, Austronesian) such as *k-ni-akri* 'act of crying', *d-i-avra* 'act of cutting', *d-ni-ivri* 'act of smashing' (high vowels become glides before a non-high vowel and delete before a high vowel; /n/ deletes after /d/; hence: [knja.kri], [dja.vra], [di.vri], with the infix completely obliterated in the last case—thus, more complex onsets, but fewer syllables than there would be in sequentially faithful pronunciations of morphological representations: ni-kakri [ni.ka.kri], i-davra [i.da.vra], ni-divri [ni.di.vri]).[10]

(ii) Syllable groupings, i.e., METRICAL FEET, with the arrangement of syllables aiming at the foot type preferred in the language.

Examples:

- construct state forms of nouns in Ulwa (Sumu, Misumalpan) such as *suu-ka-lu* '(his) dog', *siwa-ka-nak* '(his) root', *karas-ka-mak* '(his) knee' (Green 1999: Section 3.2);
- 3rd person plural subject forms of predicates in Samoan (Malayo-Polynesian, Austronesian), expressed through stem-internal CV-reduplication, as in

[9] In Latin, it is only at the STEM level that the nasal stem formative in suffixal position would yield an impermissible coda cluster, with a more sonorous followed by a less sonorous consonant (*fud-N-*, *vik-N-*, *rup-N-*); as all inflectional suffixes begin with a vowel, WORD FORMS would be syllabified acceptably without metathesis (*fud.nō*, *vik.nō*, *rup.nō*, etc.). So, phonological infixation has to be seen as either applying at stem level, motivated by a coda cluster constraint, or at word-form level, in which case the motivation would be a cross-syllable dispreference of non-sonorous coda followed by sonorous onset.

[10] See Blevins 1999, also for the complex allomorphy of the adfix/infix/parafix. Blevins's own point is that Leti provides counterevidence to the theory of infixation as phonology. However, as phonological motivation she only recognises optimisation yielding CV syllables. But it is syllabic compactness, in accordance with the phonotactics of permissible onset clusters, which is the motive for infixation in Leti—and this is also phonology.

ma-lo-losi 'they are strong', *sa-va-vali* 'they travel', *ata-ma-ma/i* 'they are clever' (Broselow & McCarthy 1983/84);
- English expletive and *-ma-* "infixation", not otherwise very typical, as in *abso-bloomin-lutely*, *secre-ma-tary* (McCarthy 1982, Yu 2004, 2007).

(iii) Consonantal patterns in STEM TEMPLATES.

Example:

- causative verb stems in Tiene (Bantu, Niger-Congo) such as *l-as-ab-* 'cause to walk', with derived verb stems in Tiene being of the shape $C_1VC_2VC_3$ and required to form a "prosodic trough" with C_2 coronal and C_3 non-coronal (Hyman 2006).

That affixes can only ever be adfixes in morphological representations (other than perhaps ones containing bipartite stems trapping former adfixes), as per the universal assumed here, is reflected by "infixes" always remaining EDGE-BOUND in pronunciation: they are never found further inside stems than after/before the initial/final constituents of the relevant prosodic unit—after any syllable-onset in Tagalog (perhaps sometimes vacillating between after the first consonant or after the entire onset cluster); after syllable-onsets in Leti, provided they yield a permissible cluster and the segmental environment permits syllabic reduction; before plosive syllable-coda in Latin; after the first iambic foot in Ulwa; before the word-final trochaic foot in Samoan; at left or right edges of final/initial trochees in English; before the final non-coronal consonant of the template in Tiene. Where infixation is specifically prominence-driven, with adfixes attaching to prosodically prominent units in phonological representations (stressed vowels or syllables, heads of feet), it is edge-bound, too, since prominence itself is determined from word-edges.

Lending further support to the phonological theory of infixation, "infixes" are, in the relevant languages, always also realised as adfixes with stems where the prosody is satisfactory without phonological rearrangement. Thus, in Tagalog, the prefix *um-* remains in place with vowel-initial stems such as *um-awit* PERF of 'sing'. In Latin, the nasal suffix remains in place with stems of the same conjugation class without a stem-final plosive such as *si-N-* 'leave', *ker-N-*'separate', *(con-)tem-N-* 'despise', *pell-* (< *pel-N-*) 'expell'. In Leti, the nominalising prefix *ni(a)-* remains in place with stems where syllables would not be compacted: *nia-keni* 'the act of placing' [nja.ke.ni]/**k-nia-eni* [knja.e.ni], *(n)i-atu* 'knowledge' [(n)ja.tu]/**a-(n)i-tu* [a.(n)i.tu]. In Ulwa, the construct suffix remains in place with iambic stems of two moras, contributed by one heavy syllable or two light ones, or of three moras, contributed by a light syllable followed by a heavy one, such as *kii-ka* '(his) stone', *sana-ka* '(his) bee', *sapaa-ka* '(his) forehead'. In Samoan, CV-reduplication is external when the final trochee is all the predicate consists of: *pe-pese* 'they sing', *la-laga* 'they weave'. In Tiene, verb stems with a final coronal form a "prosodic trough" when suffixes with a non-coronal consonant remain external, such as

mat-is- (→ *maas-*) 'cause to go away'.[11] Infixes due to entrapment are also prone to have stem-external realisations; but here the internal/external distribution tends to be regulated morphologically or indeed lexically (in terms of stem classes) rather than phonologically. All the same, there is evidence—e.g., from Lakhota (Siouan), whose verbs can have person markers prefixed or infixed depending on stem class (Albright 2000)—that even here distributions are prone to be increasingly regulated along phonological lines, corroding the contrast between origins by entrapment and metathesis.

Reasserting their morphological adfix status, as continuingly encoded in morphological representations, "infixes" can again be RE-EXTERNALISED, even from the very stems they used to surface inside of. Thus, the adfix *um-*, which is widespread in Malayo-Polynesian and whose manifest position is either before stems when stems are vowel-initial or after the first consonant of consonant-initial stems, as illustrated above from Tagalog, is increasingly found in prefix position also with consonant-initial stems in Toba-Batak, with its own final consonant assimilating in place: e.g., *ud-dátu*, now alternating with older *d-um-átu* 'wiser', *ug-gógo*, now alternating with older *g-um-ógo* 'stronger', *ul-lógo*, now alternating with older *l-um-ógo* 'drier' (Crowhurst 1998).[12] Such re-externalisations will not happen once "infixes" have become lexicalised, part of internally unanalysed stem allomorphs, which they probably were as Latin turned into the Romance vernaculars; but then, being lexicalised, they are not constructive parts of morphological representations, either.

Other than their edge-boundedness, with "edges" defined prosodically, it is the preferred phonological shapes of "infixes" which bespeak their phonological nature. In terms of segments, infixes (other than entrapped ones) typically consist of labials, palatals, pharyngeals, laryngeals, liquids, and rhotics, and typically of only a single segment of these kinds: these segments are precisely those which are most susceptible to perceptually-driven metathesis in general (Yu 2007: 139–148). Even where the internalisation of an adfix makes little or no prosodic difference vis-à-vis an external placement, susceptibility to metathesis alone may suffice to prompt rearrangements of morphological representations—e.g., the placing of the negative suffix inside final syllables of verbs in Hua (non-Austronesian, Papua), with consistent CV structures either way: *haru-ʔa-po* 'not slip' (compare **harupo-ʔa*). It is really only making prosodic matters worse which would not be expected.

What can be assumed to be crosslinguistically invariant, then, is order in morphological representations, always and ever obeying the constraint NO INFIXES!, except perhaps as the inadvertent result of entrapment or, very rarely, of excrescence. What

---

[11] Atypically, American English *-ma-* does not occur externally, and never has, not having had an adfix as its immediate source, but supposedly originating through excrescence (Footnote 7). When external, expletives (*bloomin' absolutely*, *fuckin' Chicago*, etc.) are words (adjectives or adverbs), not adfixes.

[12] With stems beginning with a labial or nasal consonant, *um-* had always remained external in Toba Batak, with /m/ assimilating in place: e.g., *up-pásak* 'has beaten', *uŋ-ŋáli* 'colder'.

is variable is how, at any given time, speech communities rank phonological (in particular, prosodic) optimality relative to morphological faithfulness in pronouncing complex words. If complex words are not required to be prosodically optimal, or if stems and affixes have phonological shapes which, when combined, yield preferred prosodies anyhow (syllables, feet, templates), then adfixes will be realised as adfixes; otherwise they will be internalised around edges, with phonological (prosodic) constituents in an order which sounds better than arrangements faithful to the morphology. Given the historical mutability of "infixation", the circumstances which license or indeed require, or also proscribe, phonological improvements of morphology would also be expected to be variable.

While it is true to say, achronically as well as diachronically, that infixes imply adfixes, this implication as such has no status in mental grammars (and, as such, is of little theoretical interest); it is (prosodic) phonology, acting on invariably infix-less morphological representations and obeying constraints of its own, that masterminds overt variation.

7.2 Conceptual Semantics of Syntactic Construction Restructured in Context

The second example of a differential constraining of representations is about stacked attributive adjectives preceding or following a noun.[13] What are subject to separate constraints here are syntactic representations of such phrases which are responsive to conceptual and scope relations on the one hand and to information structure on the other.

When adjectives of different semantic classes are to be combined with a noun in attributive constructions, two decisions are to be made: first, whether to put the adjectives (all or some) before or after the noun; second, how to order the adjectives among each other. (Adjectives are property-concept words with a grammar of their own, distinct from those of both nouns and verbs. Property-concept words of a nominal or verbal nature, with no distinct word class of adjective, should show similar positional proclivities in the languages concerned.) In languages where the ordering is relatively rigid at phrase level, the first decision is usually clear-cut; while the second tends to be less categorical, there usually are clear preferences, as illustrated for only three semantic adjective classes in the following examples:

(6) a. English et al.
a beautiful big red ball VALUE SIZE COLOUR N

b. Bahasa Indonesia et al.
bola merah besar tjantik N COLOUR SIZE VALUE

[13] See further Plank 2007.

(7) a. Maltese et al.
ballun sabiħ kbir aħmar N VALUE SIZE COLOUR
a′. Italian et al.
una bella grande palla rossa VALUE SIZE N COLOUR
b. apparently unattested
**redd bigg beautifull ball* *COLOUR SIZE VALUE N

Where such unmarked orderings among adjectives in simple, non-coordinate, cohesive noun phrases can be established,[14] the following descriptive generalisation can be made over (6a/b), which are by far the most common orderings across languages:[15]

(i) The relative distance (position class) of semantic classes of adjectives from the noun is the same, whether the noun comes after or before the adjectives. (That is, the two most common orderings are mirror images of each other.)

A higher-level descriptive generalisation over (6a/b) is as follows, drawing on conceptual differences of words subsumed under one word class and of conceptual similarities across word classes:

(ii) The nounier a modifier, the closer to the noun.

The nouniness ranking COLOUR > SIZE > VALUE is independently motivated, on language-particular and general grounds. Well-known relevant evidence includes: the nounier modifier words actually are themselves nouns or are derived from nouns (e.g., *wooll-en* MATERIAL, *medicine* PURPOSE); they can enter a morphological relation with head nouns (compounding: e.g., *snow-ball* MATERIAL, *black-ball* COLOUR); their range of applicability to nouns of different semantic classes is narrower (e.g., *beautiful*/**red idea* VALUE/*COLOUR in construction with abstract noun). The less nouny modifier words are verbal forms (participles) or are derived from verbs (e.g., *ugly* VALUE, originally deverbal, borrowed from Old Norse *ugga* 'to dread'); they do not compound; their range of application is wider.

The nouniness feature mentioned last suggests a generalisation over the pattern at issue at an even higher level, where description metamorphoses into explanation. As modifiers are stacked, their natural stacking-order, hierarchical rather than linear, reflects the conceptual closeness or distance of modifiers from ultimate heads: less nouny property-concept modifiers prefer to take scope over nounier concepts. Thus, unmarked linear order is iconically motivated by preferences of scope construal:

---

[14] The most important proviso here is that not all languages permit the stacking of adjectives to begin with; some only do so very reluctantly. Also, there may be further factors relevant for ordering, such as the length or other phonological properties of adjectives or inherent emphasis of some adjectives (such as 'big').

[15] Such a universal is richly supported in the relevant literature, most substantially in Hetzron 1978 and Sproat & Shih 1990.

(iii) Linear closeness—VALUE SIZE COLOUR N / N COLOUR SIZE VALUE—mirrors scoping hierarchy—(VALUE (SIZE (COLOUR (N))))—as itself determined by conceptual distance.

This is the sort of fundamental principle that one would like to be able to invoke as a general constraint on the construction of wholes from meaningful parts, and in particular their arrangement. And iconicity is an undoubted major force in universally governing linear order in a wide range of syntactic domains where conceptual distance and scope are a factor.

It follows from this account that the anti-iconic ordering in (7b) should not occur, and it apparently does not. But neither should the equally anti-iconic ordering (7a), which does occur, even if not so frequently. In view of the existence of (7a), the obvious question is: Why is there no mirror image of (7a), i.e., (7b)? And more alarmingly, the question is whether a prized universal, stated at whatever level of generality, as in (i), (ii), or (iii), is invalidated by the overtly anti-iconic ordering in (7a).

The universal is rescued, as constraining not "language" or such forms-in-construction per se, but one kind of syntactic representation, and (7a)'s lack of a mirror image is explained, if Maltese et al. (with Semitic and Celtic languages as *alii* on record), instantiating the surface ordering in (7a), are analysed as being like English et al.: namely as having NPs where N is in final position. This similarity can only hold at a level of syntactic representation that is not a direct input to pronunciation—at a level where linear order is dictated by scope construal determined by conceptual proximity, only concerned with rendering conceptual meaning and unencumbered by any other expressive responsibilites. Thus, as to the relative ordering among multiple adjectives, iconicity could be assumed to rule OK everywhere and timelessly for representations at such a level, and the only variable here is whether modifiers come before nouns (6a, 7a/a′) or after (6b, 7a′).

The price to pay for an account where syntactic representations—abstract insofar as they are not the representations pronounced—are universally constrained as per (i)/(ii)/(iii) is a syntactic rule of N-fronting (7a), or half-way fronting as in (7a′), exemplified by Romance, tampering with abstract order.[16] The question that comes with it is why only a few languages front or half-way front N, while many languages leave N where it is. And yet another question needs to be addressed, namely, why there are no abstract representations in line with iconicity which end up with a counter-iconic overt order through N-BACKING—that is, with overt (7b) derived from abstract (6b). Some explanatory mileage might be gotten out of the particular directional asymmetry in this respect where grammars are variable—displacing or not displacing N; but if displacement, then only by fronting, never by backing. Ordering under the iconic supervision of conceptual semantics can apparently be interfered with as the information to be presented in context is being structured, with

[16] Following Cinque 1994 and Longobardi 1994, who took their inspiration from Romance, N-fronting has been much discussed in generative syntax, with more attention paid to technical implementation than to the typological milieu of such a rule.

the requirements of topic-comment or focus-background articulation potentially advising that scope-iconic order better be deviated from. Perhaps, as modifiers will naturally be comments, the displacement of N in NPs is to be understood as a sort of topicalisation—and, universally, topics tend to be overtly fronted, not backed.[17]

So, what appears as crosslinguistic disorderliness on the surface can be reduced to order when universals are envisaged as differentially constraining syntactic representations at different levels—one where conceptual and scopal relations are represented (constrained through (i)/(ii)/(iii)) and another where information structure is represented (subject to a prohibition against N-backing). Also, what would appear to be universally invariable here is that it is information-structural representations rather than conceptual-scopal representations that are being pronounced.

7.3 Segments in Pronunciation and in Phonological Systems

Turning to phonology and phonetics, constraints that can be suggested for segment inventories need not hold at all levels of representation. For example, the universal that all languages have voiceless stops is disconfirmed by the phonetic inventories of a very few languages—ca. 1% of the languages in the UCLA Phonological Segment Inventory Database (UPSID, comprising 451 languages), mostly Australian (Bandjalang, Mbabaram, Dyirbal, Yidiny; also North Carolina Cherokee). For phonological inventories the universal can arguably be maintained in categorical form, depending, however, on phonological analysis (which is why Hyman 2008 distinguishes such phonetic and phonological universals as "descriptive" and "analytic"): for each offending language it has to be made plausible that phonetically voiced stops are phonologically voiceless (thus, Yidiny [b, d, ɟ, ɡ] would have to be analysed as underlyingly be /p, t, c, k/, with voicing redundant, or Cherokee [d, g] as underlyingly being clusters /hd, hg/, etc.).

Divergences here may be due to feature changing as well as feature spell-out. A universal claim to the effect that CORONAL is unspecified for all segments in all positions in all words of all languages would be faulted immediately: in languages that have them in their segment inventories (all do), coronals are pronounced; that is, the relevant segments—sounds pronounced with the blade of the tongue raised from its neutral position (dental, alveolar, and probably palatal consonants, front vowels)—cannot but be specified for CORONAL at the level of pronunciation. Whether CORONAL is unspecified at any or all levels of representation other than that determining pronunciation is a live issue. There is a whole family of patterns, rules, or constraints where coronal and non-coronal segments behave asymmetrically, with the patterns of inequality always the same: in any language, either only coronal segments can do something which non-coronal segments can't, or whatever non-coronals do,

[17] As to the question of why languages differ in letting or not letting nouns be displaced to begin with, another lead that remains to be pursued typologically is that languages with N-fronting are ones where V is initial, too (as in Semitic and Celtic), at least in abstract representations. Inflectional differences in the marking for gender and number have also been claimed to be implicated; but this is factually more dubious.

corresponding coronals do, too. We believe the evidence available, plausibly analysed, strongly suggests that CORONAL underspecification at all levels is a valid universal. As an instantiation of the basic structural principle of the asymmetry of contrasts, it is a universal of such pervasive scope that discussion had better be postponed to the next and last section.

8 True Throughout

In search of universals true of ALL parts and of ALL representations of forms and constructions, one would first turn to the basic organising principles of lexicons and grammars: these could be expected, and should then be demonstrated, to make their influence felt pervasively, over all parts and all representations, rather than only locally.

As an example, we will mention asymmetry as such a basic structural principle inspiring many individual universals in phonology. No part or no representation can offend against asymmetry by having the opposites reversed.

Phonological systems are centrally defined through contrasts. What counts is not lists of "phonemes", however popular these are in typology, but the finite set of properties which define segmental contrasts—distinctive features, themselves considered universal. Phonological features make up lexical representations of morphemes; these are subject to changes in the different contexts in which they are perceived and pronounced. On the evidence of synchronic alternations, of change, of acquisition, and of perception and processing, phonological rules and constraints are universally asymmetric, just as representations of contrasts themselves are fundamentally asymmetric.

8.1 Asymmetry in Phonological Inventories

The first step in describing phonological systems is to set out the consonant and vowel inventories. This is what Pāṇini did, charting the consonants of Sanskrit and ordering them by place of articulation, aspiration, and voicing. When reciting the consonants in Sanskrit (or modern Bengali), one begins with the back of the mouth (velar articulation) and ends with the labials; that is, a consonant chart like (8) is meant to be read left to right, top to bottom.

(8) Pāṇini's structured consonant system of Sanskrit

	VOICELESS		VOICED		NASAL
	UNASPIRATE	ASPIRATE	UNASPIRATE	ASPIRATE	
VELAR	k	k^h	ɡ	$ɡ^h$	ŋ
PALATAL	c	c^h	ɟ	$ɟ^h$	ɲ
RETROFLEX	ʈ	$ʈ^h$	ɖ	$ɖ^h$	ɳ
DENTAL	t	t^h	d	d^h	n
BILABIAL	p	p^h	b	b^h	m

In this chart, with no empty cells, the plosives and nasals are symmetric.[18] But in this respect Sanskrit is very unusual: most sound inventories are asymmetric.

Again on the evidence of UPSID,[19] vowels are usually contrasted in terms of height or frontness/backness. As summarised in (9), and keeping aside actual feature descriptions, if we consider only high vowels, the most frequent vowel systems have a two-way contrast; if there is a three or four-way contrast, then vowels also contrast in rounding.

(9) Number of languages from UPSID with [i], [u], [y], [ɯ]

	FRONT		BACK	
ROUND	26	[y]	383	[u]
UNROUND	411	[i]	41	[ɯ]

The most common high vowel is front unrounded [i], followed closely by back rounded [u]. The other two are almost equally infrequent. When we compare the languages which have these combinations, we find only four languages with four vowels, while 62 languages have a three-way contrast and 376 languages have a two-way contrast. However, not all combinations are equally likely: two-way contrasts predominantly include [i] and [u]; no two-way contrast includes [y] and [ɯ]; three-way contrasts invariably include either [u] or [i], as there are 31 languages with [i u ɯ] and 19 with [i u y]; of all languages which have [u] (383), only five have no [i]; of the 411 languages with [i], 36 have no [u].

This suggest the following implicational universal: The presence of [y] or [ɯ] in a vowel system implies the presence of either [i], [u], or both. This means that if there is a two-way contrast between high vowels, either [i] or [u] must exist. Accordingly, if a system changes such that a vowel of a three-way inventory is lost or merges with another, the vowel affected will be either [y] or [ɯ]. This is precisely what happened in English. Old English had a contrast between [i], [u], and [y], while the Modern English contrast is betwen [i] and [u]. All [y]'s have become unrounded, yielding [i].

This gets us directly to the question of phonological rules and their outputs. As mentioned earlier, contrasts in lexical representations do not remain unchanged. In fact, more often than not, contrasts may be neutralised or changed in different contexts. Here, too, we see indications of pervasive asymmetry.

8.2 Asymmetry in Phonological Rules

Phonological rules can be broadly divided into four types depending on whether they (i) introduce features, (ii) neutralise contrasts, (iii) delete or (iv) add segments. Rules that delete or add segments are usually governed by structural constraints,

[18] The precise phonetic/phonological status of palatals, debated controversially, is not an issue here.

[19] Like Hyman 2008, we have used UPSID with the web interface developed by Henning Reetz: http://iona.sprachwiss.uni-konstanz.de/L/L4904.html

and those that add new and non-contrastive features are allophonic and, more often than not, phonetic in nature. Those that neutralise contrasts are the most problematic since they produce morphemic variations and alternations.

The phonological rules (both neutralising and allophonic) which most commonly ensue in featural changes are assimilation rules, with contiguous segments becoming closer in phonological features. A remarkable fact about these rules is that they are unidirectional: the reverse never occurs in corresponding contexts.

(10) Common assimilatory rules

i.	Vowel Nasalisation:	V → Ṽ /__ nasal C
	But not:	Ṽ → V /__ oral C
ii.	Umlaut or V-fronting:	/u/ → [y] /__/i, j/
	But not:	/y/ → [u] /__/u, w/
iii.	Palatalisation:	/k/ → [t∫] /__/i, j/
	But not:	/t∫/ → [t†] /__/u, w/
iv.	Retroflexion:	/t/ → [t̠] /__ high back C or V
	But not:	/t̠/ → [t] /__ low front C or V
v.	Rounding:	/e, a/ → [o, ɔ] /__/u/
	But not:	/o, ɔ/ → [e, a] /__/i/

Although such assimilation rules never operate in reverse, reverse changes can occur, but not as assimilations. For example, nasal vowels can be denasalised, but the change then is not assimilatory in nature and can happen without any context. Moreover, if the assimilated phonemes change, they do not necessarily revert back to their origin. Umlauted vowels like /y/ can become de-umlauted, but do not revert back to /u/; rather, they become unrounded /i/. Thus, feature changes are in essence asymmetric.

A further aspect of assimilation rules is the effect they have on the system: they can be allophonic or neutralising. All the rules mentioned above can be both. But again there is asymmetry, insofar as allophonic rules will become neutralising once the new feature has become contrastive, whereas neutralising rules cannot become allophonic. Consider the same rules as in (10).

(11) Allophonic to neutralising: adding new contrasts

i.	Vowel Nasalisation:	V → Ṽ / ~~__ nasal C~~
ii.	Umlaut or V-fronting:	/u/ → [y] / ~~__/i, j~~/
iii.	Palatalisation:	/k/ → [t†] / ~~__/i, j~~/
iv.	Retroflexion:	/t/ → [t̠] / ~~__ high back C or V~~
v.	Rounding:	/e, a/ → [o, ɔ] / ~~__/u~~/

What happens diachronically is that the contexts which led to the assimilations are deleted (as indicated) or are otherwise no longer transparent. A case in point is vowel nasalisation in Indo-Aryan languages. The modern Bengali descendant

of Sanskrit *panca* is [pãtʃ] 'five', with the vowel nasal and the nasal consonant deleted. This is a typical example where an allophonic rule of vowel nasalisation in the context of a nasal consonant has led to a new contrast being added. Vowels in Bengali contrast in nasality as in [kãda] 'to cry' vs. [kada] 'mud'. The rule of vowel nasalisation still exists, but now it is neutralising: [tʃa] 'want' (imperative), [tʃae] 3SG.PRESENT, [tʃãn] 3SG.HONORIFIC.PRESENT. A similar situation exists in German. Umlaut was allophonic in high vowels in Old High German, but is now contrastive in New High German: cf. OHG *fuß – füßi* 'foot' SG–PL, NHG *Fuß – Füße*; OHG *türi – türi* 'door' SG–PL, NHG *Tür – Türen*. The context of umlauting has been reduced to a schwa in the plural and has been deleted in the singular for 'door'. Consequently there is an extra vowel in the phoneme inventory for modern German, /y/.

To illustrate schematically how a phonological system may change, assume that a language has the three segments, two consonants, /p/ and /m/, and one vowel, /a/. Several rules can affect words like /pam/ or /pap/ made up of these three segments.

(12) Changes in phonological systems

Scenario A				
Representation		/pam/	/pap/	/pamp/
Change 1	vowel nasalisation	pãm	----	pãmp
Change 2	final consonant deleted	pã	pa	pãm
Output		[pã]	[pa]	[pãm]

Scenario B				
Representation		/pam/	/pap/	/pamp/
Change 1	vowel nasalisation	pãm	----	pãmp
Change 2	initial consonant deleted	ãm	ap	ãmp
Output		[ãm]	[ap]	[ãmp]

The two scenarios differ with respect to which consonant is deleted. Vowel nasalisation in both scenarios initially leads to an allophonic alternation: [a] before an oral consonant, [ã] before a nasal consonant. However, after consonant deletion, we have a new contrast in Scenario A since vowel nasalisation is now no longer predictable: both [a] and [ã] occur at the end of a word, and the latter also occurs before nasal consonants. Scenario B is different. Here, despite consonant deletion, the nasality of the vowel remains allophonic: it always occurs before a nasal consonant. In Scenario A, not only does the nasalised vowel become contrastive after final consonant deletion, but the rule of vowel nasalisation still functions before nasal consonants. Thus, nasality on vowels becomes contrastive AND vowel nasalisation consequently becomes a neutralising rule rather than allophonic. In contrast, nothing has changed in Scenario B: no new contrast has been added and the allophonic status of vowel nasalisation remains unaltered.

The gist of this schematic example is not that a new "sound" has been added in Scenario A, but rather that a feature has become contrastive—nasalisation on vowels. It is features rather than phonemes which are at the heart of phonology, and therefore also of phonological universals, and both contrasts and phonological alternations are governed by the specification of features. What we require of a

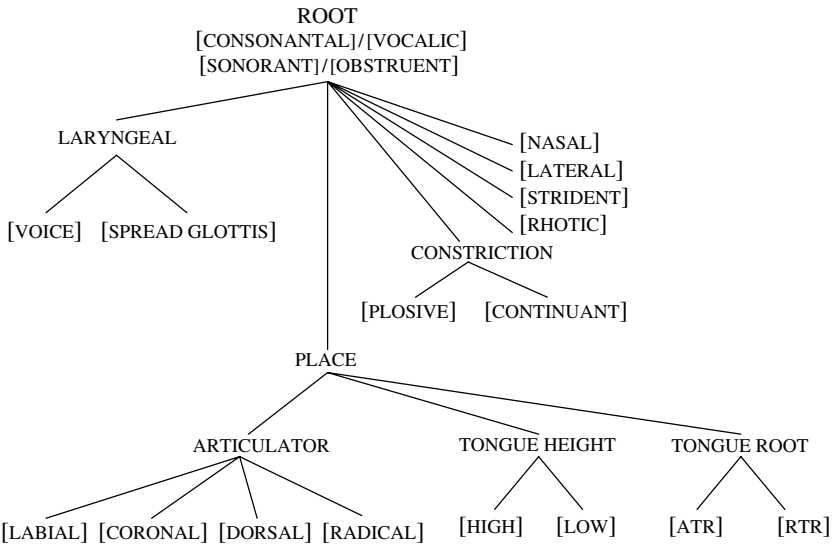
phonological feature system is that it should be able to account for universal contrasts as well as for asymmetries in the output of phonological rules and for constraints on phonological change. As a pervasive organising principle of the grammar of sounds, the same feature system should also be able to deal with language production and comprehension. Asymmetry is an essential structural principle of the feature system which we think responds to these challenges well, as to be outlined now.[20]

8.3 Asymmetry and Phonological Feature Organisation

Asymmetry is inherent to features, feature distributions, and the direction of phonological rules. The question is how to account for this in modeling mental lexicons-and-grammars. One approach among several in the literature is underspecification (Dresher & Zhang 2007, Ghini 2001, Kabak 2007, Lahiri & Reetz 2002), holding that asymmetries are encoded directly through the lack of featural specification. Objecting to underspecification, proponents of full specification have devised a variety of extra mechanisms to get asymmetry grafted onto symmetric contrasts, or they recognise some degree of underspecification for purposes such as allophonic alternations; but here is not the opportunity to really argue for the superiority of the "direct" approach that we adopt (see Lahiri & Reetz 2007).

For the feature theory of our underspecification approach, FUL (short for Featurally Underspecified Lexicon), two crucial assumptions are that consonants and vowels share features (cf. also Clements & Hume 1995) and that place features are subdivided into ARTICULATOR and TONGUE HEIGHT or APERTURE features.

(13) FUL's feature system



[20] Based on earlier work by Lahiri & Reetz 2002. Also see Ghini 2001, Lahiri & Evers 1991, and a more detailed description in Lahiri & Reetz 2007.

The feature tree in (13) embodies several claims about universal contrasts. Lexical representations in all languages must distinguish CONSONANT and VOWEL. A further universal contrast is that between SONORANT and OBSTRUENT. Furthermore, no language will lack a contrast in PLACE of articulation. Since PLACE is divided into ARTICULATOR and TONGUE HEIGHT, the claim is that all languages must have at least two consonants contrasting in ARTICULATOR. And here comes the first role of underspecification: the consonantal contrast will universally be between underspecified CORONAL and another ARTICULATOR feature. Vowels, however, need not contrast in ARTICULATOR, and theoretically a language could have only one vowel. In this case there would be no PLACE feature necessary. If a contrast in vowels exists, then the first cut will universally be in TONGUE HEIGHT, contrasting HIGH and LOW.[21] While every language will have a PLACE contrast in consonants, it will not necessarily have a LARYNGEAL contrast. Further, there are no feature dependencies: STRIDENT can be a property of non-coronal consonants.

Inevitably following from this conception, assimilation rules which spread features are asymmetric. Thus, PLACE assimilation is predicted to spread specified place features (LABIAL, DORSAL) to unspecified CORONAL. This does not mean that LABIAL consonants cannot become DORSAL in a similar context; but this would imply allomorphy on the part of morphemes involved in such a process. Further, a rule of place assimilation for labials (ii) cannot exist without an assimilation rule for coronals (i):[22]

(14) Asymmetric PLACE assimilation

i. /n/ /b/ → [mb]
unspecified LABIAL
single morpheme with /nb/

ii. /m/ /g/ → [ŋg]
LABIAL DORSAL
implies allomorphy /mg/–/ŋg/

Underspecification does not depend on syllable structure nor on possible allophonic processes alone. For example, while it is often assumed that for word-final place assimilation final coronals are underspecified because they are vulnerable to change, in our approach CORONAL is underspecified in all positions, initial as well as medial, too.

---

[21] For Dresher & Zhang 2007 languages can differ with respect to which feature is underspecified. Consequently CORONAL need not always be underspecified; if CORONAL is active in a phonological process, it will be specified. In the version of underspecification that we assume (without really arguing for it on this occasion) no such crosslinguistic variability is envisaged: underspecification is not subject to any activity — it is a representational fact.

[22] See also Mohanan 1993, as confirmed for a wider range of languages in Jun 2004.

8.4 Feature Asymmetry in Perception

The pervasiveness of featural asymmetry also extends to speech perception, and we conclude by briefly summarising relevant experimental evidence.

The assumption is that variation in speech is resolved by the listener in two steps: (i) the auditory system parses the acoustic signal into features and not segments; (ii) a mapping process, using a ternary logic of *match*, *mismatch* and *nomismatch*, matches the features extracted from the acoustic signal with those stored in the mental lexicon.

The *match* condition is transparent. A *mismatch* results when a feature extracted from the signal is in conflict with the feature in the representation. However, certain non-perfect matches are tolerated due to underspecification: this is the *nomismatch* condition. Matching predictions for consonants are given in (15), with no feature within brackets indicating underspecification.

(15) Mapping of features for consonants

Signal		Match	Representation	
[p, b, m]	LAB	*NOMISMATCH*	/t, d, n/	[ ]
[t, d, n]	COR	*MISMATCH*	/p, b, m/	LAB
[k, ɡ, ŋ]	DOR	*MISMATCH*	/p, b, m/	LAB
[t, d]	COR	*MISMATCH*	/k, ɡ/	DOR
[k, ɡ, ŋ]	DOR	*NOMISMATCH*	/t, d, n/	[ ]

Through a semantic priming task (lexical decision, crossmodal) we tested CORONAL underspecification in word-medial and word-final positions in German (Lahiri & Reetz 2002). For the medial condition, where no assimilation is ever possible, words like *Ho[n]ig* 'honey' predictably facilitated recognition of *Biene* 'bee', and *Ha[m]er* 'hammer' primed *Nagel* 'nail'. Pseudoword variants of these primes, however, gave asymmetric results: **Ho[m]ig* successfully primed *Biene*, but **Ha[n]er* did not prime *Nagel*. That is, the LABIAL [m] of the pseudoword **Ho[m]ig* was tolerated as a variant of the underspecified /n/ in *Ho[n]ig* and sucessfully facilitated the recognition of *Biene*; but the coronal [n] of the pseudoword **Ha[n]er* was rejected by the lexically specified /m/ in *Ha[m]er*.

In a more recent electro-encephalographic (EEG) study using words varying in medial coronal vs. non-coronal consonants we examined whether such an asymmetry would also be found with a more direct technique for measuring brain activity (Friedrich et al. 2006). Word-medial coronals in *Hor[d]e* 'horde' are placeless in our feature theory, and the claim is that their corresponding non-coronal variant, as in **Hor[b]e*, cannot mismatch this empty PLACE slot and therefore would activate *Hor[d]e*. A similar mapping would not occur with pseudowords with a coronal like **Pro[d]e* and a corresponding real word *Pro[b]e* 'test'. CORONAL extracted from **Pro[d]e* mismatches the specified LABIAL of *Pro[b]e* and therefore cannot activate this word. The prediction is that lexico-semantic memory search processes would be successful when **Hor[b]e* is presented and activates the corresponding coronal word *Hor[d]e*, but not when the coronal variant **Pro[d]e* is presented, since this

would lead to an immediate correct rejection as a non-existing lexical item. Thus, an asymmetry was expected at least for the initial N400 pseudoword effect, which is most likely related to lexico-semantic processing.

The task was speeded lexical decision to auditory stimuli. For the behavioural results, the error rates revealed significant differences. Non-coronal pseudowords like **Hor[b]e* (*<Hor[d]e*) had significantly more errors than coronal pseudowords like **Prode* (*<Pro[b]e*), suggesting that subjects more easily recognised **Pro[d]e* as a nonword, but had more difficulty in rejecting **Hor[b]e* as a nonword since it did activate the real word *Hor[d]e*. In the ERP data, the early N400 results showed a clear asymmetry in the earlier activation period of 100–250 ms. Mean amplitudes of the coronal pseudoword variants were significantly more negative than their non-coronal base words. By contrast, ERPs for non-coronal variants did not differ from their base words in this initial part of the N400 pseudoword effect. Furthermore, a significant difference between both types of pseudoword variants, but not between both types of words, relates this early ERP deflection to mismatch detection in the case of coronal pseudowords.

Thus, medial coronal consonants, which contrast with dorsal and labial consonants, also show an asymmetric pattern. Non-coronal pseudowords with labial or dorsal consonants are accepted as variants of the corresponding coronal word, but not vice versa as shown in the error data as well as in the early N400 effect. Medial consonants do not undergo any assimilation such that the pseudowords could have been "experienced" or become familiar to the listeners. Further, since word frequency was controlled, full specification or specification of phonetic detail cannot account for these results.

The same predictions hold for vowels. Given the underlying contrast between /o e ø/ in German, /o/ is specified for DORSAL and LABIAL, while the others are unspecified for CORONAL.

(16) Underspecified vowel representation

Signal		Match	Representation	
[o]	DOR	*NOMISMATCH*	/ø/	[ ] LAB
		NOMISMATCH	/e/	[ ] [ ]
[e], [ø]	COR	*NOMISMATCH*	/ø/	[ ] LAB
			/e/	[ ] [ ]
[e], [ø]	COR	*NOMISMATCH*	/o/	DOR, LAB

A magneto-encephalographic study reports topographic differences in the processing of mutually exclusive isolated CORONAL and DORSAL vowels in German (Obleser et al. 2004). Further, Eulitz & Lahiri (2004) used a component of the event-related brain activity, Mismatch Negativity (MMN), to investigate the issue of asymmetry in mapping. MMN is assumed to be an automatic detection measure of the brain's ability to detect change in sounds, particularly to phonemes. If a sound is presented many times in a sequence (known as the "standard"), it is considered to tap the long-term sound representation, that is, our underlying lexical representation. If another sound is presented right after the sequence (a "deviant"),

it would cause something of a jolt, and the brain would detect a change and respond accordingly. The classical MMN is a high-amplitude difference between standard and deviant around 180 ms from the onset. Eulitz & Lahiri (2004) noted both an amplitude and a latency difference. As predicted by the matching algorithm, for the pair [o]~[ø], when /o/ was the standard (i.e., underlyingly specified for DORSAL) and [ø] the deviant such that [CORONAL] is extracted, there was a higher and earlier MMN peak than the other way round. Similar predictably asymmetric patterns of results were obtained for the other pairs. Thus, just as for the consonants, the vowels showed asymmetric perceptual responses as predicted by our approach to featural underspecification.

With such confirming experimental evidence from only a few languages so far, we are nevertheless confident that at no level of representation—structuring how words are stored in the mental lexicon, how they are accessed, and how they are perceived and pronounced—can featural contrasts, and the phonological systems and rules defined through them, EVER be at odds with asymmetry.

Acknowledgments We are grateful to the Sonderforschungsbereich 471 at Konstanz, funded by the Deutsche Forschungsgemeinschaft, for encouraging joint undertakings like ours, bringing about the present paper. Authors' names are in alphabetical order; the phonology is (mostly) Lahiri's, the morphology and syntax (mostly) Plank's, and our typological conclusions emerged and connected along the way. Sergio Scalise was a most supportive editor, and we'd also like to thank one anonymous reviewer for critical, but sympathetic comments and questions.

References

Albright, Adam. 2000. The productivity of infixation in Lakhota. *UCLA Working Papers in Linguistics* (forthcoming).

Blevins, Juliette. 1999. Untangling Leti infixation. *Oceanic Linguistics* 38: 383–403.

Broselow, Ellen and John J. McCarthy. 1983/84. A theory of internal reduplication. *The Linguistic Review* 3: 25–88.

Cinque, Guglielmo. 1994. Evidence for partial N-movement in the Romance DP. In *Paths towards Universal Grammar: Studies in Honor of Richard S. Kayne*, eds. Guglielmo Cinque et al., 85–110. Washington, D.C.: Georgetown University Press.

Clements, George N. and Elisabeth V. Hume. 1995. The internal organization of speech sounds. In *The Handbook of Phonological Theory*, ed. John A. Goldsmith, 245–306. Oxford: Blackwell.

Crowhurst, Megan J. 1998. *Um* infixation and prefixation in Toba Batak. *Language* 74: 590–604.

Dresher, B. Elan and Xi Zhang. 2007. Contrast and phonological activity in Manchu vowel systems. *Canadian Journal of Linguistics* 50: 45–82.

Eulitz, Carsten and Aditi Lahiri. 2004. Neurobiological evidence for abstract phonological representations in the mental lexicon during speech recognition. *Journal of Cognitive Neuroscience* 16: 577–583.

Friedrich, Claudia, Carsten Eulitz, and Aditi Lahiri. 2006. Not every pseudoword disrupts word recognition: An ERP study. *Behavioral and Brain Functions* 2: 1–36. (http://www.behavioralandbrainfunctions.com/content/2/1/36)

Ghini, Mirco. 2001. *Asymmetries in the Phonology of Miogliola*. Berlin: Mouton de Gruyter.

Green, Thomas Michael. 1999. A lexicographic study of Ulwa. Doctoral dissertation, MIT. (http://www.slaxicon.org/papers/thesis.pdf)

Hetzron, Robert. 1978. On the relative order of adjectives. In *Language Universals*, ed. Hansjakob Seiler, 165–184. Tübingen: Narr.

Hyman, Larry M. 2006. Affixation by place of articulation: Rare and mysterious. Paper at the conference on *Rara*, Leipzig, 29 March – 1 April.

Hyman, Larry M. 2008. Universals in phonology. *The Linguistic Review* 25: 83–138.

Jun, Jongho. 2004. Place assimilation. In *Phonetically Based Phonology*, eds. Bruce Hayes, Robert Kirchner, and Donca Steriade, Cambridge: 58–86. Cambridge University Press.

Kabak, Barış. 2007. Hiatus resolution in Turkish: An underspecification account. *Lingua*, Steriade, Cambridge: 58–86.

Lahiri, Aditi and Vincent Evers. 1991. Palatalization and coronality. In *The Special Status of Coronals: Internal and External Evidence*, eds. Carol Paradis and Jean-François Prunet, 79–100. San Diego: Academic Press.

Lahiri, Aditi and Henning Reetz. 2002. Underspecified recognition. In *Laboratory Phonology 7*, eds. Carlos Gussenhoven and Natasha Warner, 637–674. Berlin: Mouton de Gruyter.

Lahiri, Aditi and Henning Reetz. 2007. Distinctive features: Signal to representation. In *Phonetic Bases for Distinctive Features*, ed. George N. Clements (forthcoming).

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25: 609–665.

McCarthy, John J. 1982. Prosodic structure and expletive infixation. *Language* 58: 574–590.

Mohanan, K. P. 1993. Fields of attraction in phonology. In *The Last Phonological Rule*, ed. John Goldsmith, 61–116. Chicago: University of Chicago Press.

Moravcsik, Edith A. 1977. *On rules of infixing*. Bloomington: Indiana University Linguistics Club.

Moravcsik, Edith A. 2000. Infixation. In *Morphology: An International Handbook on Inflection and Word-Formation*, eds. Geert Booij, Christian Lehmann, and Joachim Mugdan, vol. 1, 545–552. Berlin: Walter de Gruyter.

Newmeyer, Frederick J. 2008. Universals in syntax. *The Linguistic Review* 25:35–82.

Obleser, Jonas, Aditi Lahiri, and Carsten Eulitz. 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. *Journal of Cognitive Neuroscience* 16: 31–39.

Plank, Frans. 1989. On Humboldt on the dual. In *Linguistic Categorization*, eds. Roberta Corrigan, Fred Eckman, and Michael Noonan, 293–333. Amsterdam: Benjamins.

Plank, Frans. 2003. There's more than one way to make sense of one-way implications – and sense they need to be made of. *Linguistic Typology* 7: 128–139.

Plank, Frans. 2007. Extent and limits of linguistic diversity as the remit of typology – but through constraints on WHAT is diversity limited? *Linguistic Typology* 11: 43–68.

Plank, Frans, and Wolfgang Schellinger. 1997. The uneven distribution of genders over numbers: Greenberg Nos. 37 and 45. *Linguistic Typology* 1: 53–101.

Sproat, Richard and Chilin Shih. 1990. The cross-linguistic distribution of adjective ordering restrictions. In *Interdisciplinary Approaches to Language: Essays in Honor of Sige-Yuki Kuroda*, eds. Carol Georgopoulos and Roberta Ishihara, 565–593. Dordrecht: Kluwer.

Ultan, Russell. 1975. Infixes and their origins. In *Linguistic Workshop III*, ed. Hansjakob Seiler, 157–205. München: Fink.

Yu, Alan C. L. 2004. Reduplication in English Homeric infixation. *NELS* 34: 619–633.

Yu, Alan C. L. 2007. *A Natural History of Infixation*. Oxford: Oxford University Press.

Universals of Prosodic Structure

Irene Vogel

Abstract The systematic study of prosodic structures is relatively recent, and the study of universals of these structures is even more recent. This paper presents a total of ten possible universals, some of which have been extracted from the original presentation of prosodic phonology in Nespor and Vogel (1986). Additional universals have been proposed based on further examination of the model as well as certain proposed modifications, in particular with relation to the Strict Layer Hypothesis. The universals are presented and analyzed in relation to three categories: (a) general properties of prosodic structure, (b) the geometry of prosodic constituents, and (c) prosodic structure phenomena. Several areas in which potential universals have been challenged are examined, and the more general question is discussed of what it means for a theory if apparent counter-examples are presented.

Keywords Prosodic phonology · phonological constituents · non-isomorphism · strict layer hypothesis · recursive constituents

1 Introduction

While we are able to look at universals and typology in many areas of linguistics with 40 years of perspective, this is not the case with respect to prosodic phonology. We might trace the beginning of this theory back to Selkirk's (1972) study of French in which she showed that certain aspects of phonology require domains not found elsewhere in grammar, or even further back to Panini's insight that there are certain phonological rules that apply in relation to morpheme and word junctures. From the perspective of universals, however, the investigation can only really begin about 20 years ago, when a full model of prosodic phonology was developed (cf. among others Selkirk 1980, 1981, 1986; Nespor and Vogel 1982, 1983, 1986; Hayes 1989).

The study of linguistic universals typically targets two general areas: abstract structures and surface patterns. To some extent, this distinction has often correlated

I. Vogel
University of Delaware

S. Scalise et al. (eds.), *Universals of Language Today*, 59–82,
DOI: 10.1007/978-1-4020-8825-4_4,

with the tendency of syntactic studies to focus in depth on a single language, while phonological studies frequently involve reference to several or many languages. As Comrie (1981), points out, we observe this distinction in the comparison of two of the major theoretical works of early generative grammar, despite their titles. *Aspects of the Theory of Syntax* (Chomsky 1965) would appear to be a general cross-linguistic study, but in fact, essentially only deals with English syntax. By contrast, *The Sound Pattern of English* (Chomsky and Halle 1968) would appear to be a study of English phonology, however, it includes reference to over 90 languages.

The present investigation of universals of prosodic phonology falls to some extent into both categories, addressing issues of abstract representation as well as the appearance of surface patterns. In the roughly twenty years of the development of prosodic phonology, numerous languages have been studied in relation to different aspects of the theory. Two issues that have received substantial attention from the perspective of universals are the inventory of prosodic constituents and the principle responsible for their hierarchical organization, the Strict Layer Hypothesis. These issues will be addressed below, along with several other potential universals based on the model of prosodic phonology. In evaluating the proposed universals, we will also consider the general question of how the observance or lack of observance of a universal affects the fundamental nature of the theory itself. That is, we must ask whether an apparent contradiction of a proposed universal means that it must be rejected completely or whether it might nevertheless represent a general linguistic tendency. We must also consider whether a reformulation of the universal is possible that better captures the fact.

In the following sections, the core aspects of prosodic phonology will first be outlined, as these will serve as the basis for the discussion of potential universals. Subsequently, a series of potential universals will be presented and evaluated. They will be grouped into the following categories: general properties of prosodic structure, the geometry of prosodic constituents, and prosodic structure phenomena.

2 Prosodic Phonology: Background

The term "prosodic phonology" is ambiguous in that it may refer either to (a) phonological phenomena termed "prosodic", as opposed to segmental (e.g. tone, stress, intonation), or (b) the theory of phonology consisting of a hierarchically arranged set of constituents that are not necessarily isomorphic to constituents of other components of grammar. We will be concerned with the latter here, although the investigation of prosodic constituent structures often involves the application of prosodic (i.e. non-segmental) phenomena.

2.1 Prosodic Constituent Structure

The prosodic constituent hierarchy is widely taken to include the items shown in (1) (cf. among others Nespor and Vogel 1986, henceforth N&V). In this structure

segments, which are not themselves prosodic constituents, are grouped into moras, the smallest prosodic constituents. Through the level of the Foot, the constituents may be considered purely phonological, while those at higher levels involve an interface with one or more other components of grammar, as indicated.

(1) Prosodic Constituents

Constituent	*Interface*
Phonological Utterance	(syntax, discourse)
Intonational Phrase	(syntax, semantics)
Phonological Phrase	(syntax)
Clitic Group	(morphology, syntax)
Phonological Word	(morphology)

Foot	
Syllable	
Mora	
(Segment)	

Different analyses vary somewhat with respect to which constituents are included in the hierarchy, as will be discussed below, but it is generally accepted that at least some set of prosodic constituents is needed in phonology.

2.2 Role of Prosodic Constituents

Prosodic constituents crucially delimit the domains of phonological phenomena that extend beyond segments, Panini's internal and external Sandhi phenomena. Any type of phonological phenomenon appears to be a potential candidate for applying in relation to prosodic constituents (e.g. phonotactic restrictions, segmental rules, non-segmental rules). It is generally assumed, following Selkirk's original analysis of French (1972), that the domains delimited by prosodic constituents are not necessarily isomorphic to other linguistic constituents. Specifically, Selkirk showed that the domain for Liaison could not be a syntactic constituent, since it applies in cases such as (2a) but not (2b), both of which are typical NPs of French. Instead, a distinct constituent structure is needed as indicated with the parentheses in (3).

(2) French Liaison

a. Le savan [t]anglais	b. Le savant // anglais
the wise Englishman	the scholar English
'the wise Englishman'	'the English scholar'

(3) Syntactic vs. Phonological Constituents

a. [le savant anglais]$_{NP}$	b. [le savant anglais]$_{NP}$
( )	(( ) . . .)

Numerous additional cases of non-isomorphism can be found across languages, and are frequently referred to as "bracketing paradoxes". For example, in the

well-known case of *ungrammaticality*, the morphology requires the prefix *un-* to be added before the suffix *-ity*, while phonologically the opposite appears to be the case. It was precisely on the basis of such cases of non-isomorphism between phonological and other structures that the theory of prosodic phonology was further developed.

3 Universals: General Properties of Prosodic Structure

3.1 Establishment of the Prosodic Hierarchy

If we assume that the fundamental organizing principles of language are part of Universal Grammar (UG), and thus a component of all languages, the first universal below is relatively uncontroversial. That is, it is essentially a definitional property of any model of prosodic phonology. The second one, however, involves an aspect of prosodic phonology that has engendered some controversy with regard to the universality of the theory, as will be discussed.

Taking as a starting point a basic model of prosodic phonology, we may posit a universal such as the one in (4). In fact, in the numerous studies of a broad range of languages that have been conducted in relation to prosodic phonology, it is widely accepted that that prosodic constituents cannot be dispensed with completely.

(4) **Universal 1: Prosodic Constituents are found in all languages.**

As mentioned above, these constituents are not necessarily isomorphic to other grammatical constituents, but rather are constructed via a mapping from morphological, syntactic, semantic and discourse structures.

It should be noted that even if we find a language where all phonological phenomena apply in relation to constituents found in other components of grammar, this would not necessarily invalidate Universal 1. It might mean that we are dealing with a universal tendency rather than an absolute universal. It might also turn out that this hypothetical type of language has some other property that predicts its phonological properties. Thus far, however, analyses of numerous languages all appear to find at least some phenomena for which the possibility of non-isomorphism is crucial.

The most obvious aspect of prosodic phonology to examine from the perspective of its universal properties is the inventory of constituents, and indeed, this has been the subject of much of the research attempting to assess the theory's universality. Following N&V, the strongest claim we can advance is Universal 2, according to which all languages include the same set of prosodic constituents:[1]

(5) **Universal 2: All languages include the following Prosodic Constituents Prosodic Word, Clitic Group, Phonological Phrase, Intonational Phrase, Phonological Utterance**

[1] These are the constituents in N&V; however, it will be shown below that they are not necessarily still defined in the same ways.

It is frequently difficult to establish the domain of application of phonological phenomena since most descriptions of languages do not provide a complete enough set of data, especially in the case of larger domains. Thus, much of the work investigating the universality of the constituent hierarchy has targeted the Phonological Word level, as seen in the next section.

3.2 The Status of the Phonological Word (PW)

The Phonological Word has undergone more scrutiny than other constituents in the prosodic hierarchy for at least two reasons. First, there is generally at least some material in phonological descriptions of languages regarding the application of phenomena at various morphological junctures. In addition, since such phenomena apply to word-sized structures, their investigation requires much less additional construction and testing of varied and possibly complex syntactic constructions.

In fact, in recent years several substantial works have been devoted to the examination of the potential universality of the Phonological Word (PW). The general consensus appears to be that the PW is a necessary, universal, component of the prosodic hierarchy, as evidenced in the collections of papers on an impressive variety of languages in Hall and Kleinhenz (1999) and Dixon and Alexandra (2002). There are differences with regard to precisely what may be included in the PW, such as different types of affixes, clitics and other types of function words. Regardless of the particular choices made in this regard, however, the PW itself is generally considered crucial.

One exception appears in the work of Bickel and Hildebrandt (2007) and Schiering et al. (2007) where it is argued that certain languages, for example, Vietnamese, lack the Phonological Word since, it is claimed, no positive evidence has been found for this constituent. Given the difficulty of investigating all the possibly subtle manifestations of prosodic phenomena, one must exercise caution in drawing the conclusion that a particular constituent is absent from a language. It is difficult to be sure that all possibilities have been considered. Even if it could be positively determined that there was no surface manifestation of the PW, or some other constituent, in a language, this would not necessarily mean that the language lacks the constituent at some deeper level. In a typical view of UG, abstract constructs are posited for human language in general, even if a language lacks an overt manifestation of a particular construct, for example, tense markers in Chinese.

From the opposite perspective, it has also been proposed that the PW needs to be supplemented. Downing (1999), for example, argues that an additional, morphologically defined, constituent is needed for Bantu languages, the Prosodic Stem. As Booij (2007) shows, however, the existence of overlap between certain morphological and phonological structures does not require the conclusion that the morphological structures are themselves phonological constituents. While it is clear that a simple analysis of the Bantu phenomena in question is not available, we must consider the consequences of introducing a single constituent that is isomorphic, by definition, to a constituent of another component of grammar. In such

a case, it would be necessary to identify two distinct categories of constituents, those that must be isomorphic to other constituents of grammar and those that are not required to be isomorphic to any other constituents. This, in turn, would require that we establish independent criteria for distinguishing the two categories of constituents.

A more extreme position is taken by Bickel and Hildebrandt (2007) and Schiering et al. (2007) on the basis of their survey of approximately 60 languages. Instead of a single Phonological Word constituent, it is proposed that languages may have any number of PWs. For example, it is claimed that Limbu has 4 PWs, while Belhare and Kham have 12 and 13 PWs, respectively. Again, it is clear that the analysis of the languages under investigation is far from simple or straightforward, thus one must exercise caution both in interpreting the data and in assessing the implications of such claims for any model of grammar.

First, it should be noted that in the cases in which languages are claimed to have particularly large numbers of Phonological Words, this is generally the result of assigning distinct domains for each of a host of phonological phenomena. Such an approach is reminiscent of earlier works in generative phonology in which numerous affixes in certain languages were assigned different boundary symbols to account for their slightly different behavior with respect to other affixes. In some cases, this led to an extensive proliferation of boundary types (e.g. Danish: 5 (Basbøll 1975, 1981); Dakota: 11 (Carter 1974)). It is not clear, however, that the appropriate way to account for individual, possibly exceptional, phonological patterns is via the establishment of different boundary symbols, or different prosodic constituents. Instead, it might be more insightful to formulate the phenomena in question in a slightly different way so as to account for the exceptionality, as suggested in the next section. Alternatively, we might simply mark idiosyncratic phonological behaviors directly on the relevant items without modifying the fundamental structure of the phonological component of grammar.[2]

Second, it is important to consider the consequences of analyses with unlimited numbers of Phonological Words, or other constituents, for a theory of phonology, and ultimately of grammar in general. If one of the goals of linguistic theory is to identify the properties of human language, we are no longer able to achieve such a goal if there is no restriction on the number of constituents involved in organizing phonological structures, the composition of these constituents, and the criteria employed for establishing the constituents of a language. Simply positing different structures for different languages in the absence of any underlying motivation for such differences is tantamount to saying that phonological systems are free to vary in essentially unconstrained ways. Furthermore, if any, unconstrained, type of analysis is considered a manifestation of a possible phonological system, there can be no way to evaluate the validity or appropriateness of a given analysis.

[2] See Simon and Wiese (to appear) for detailed discussion of different means of addressing phonological exceptions.

3.3 The Status and Properties of the Clitic Group (CG)

As just seen, it has only occasionally been suggested that the Phonological Word may not be needed as a constituent of the prosodic hierarchy. By contrast, the Clitic Group (CG) has more frequently been considered a superfluous constituent. Most of the challenges to the CG reflect problems with the application of the Strict Layer Hypothesis, according to which constituents at each level of the prosodic hierarchy must dominate constituents of the directly lower level (cf. Hall 1999).

In fact, the Strict Layer Hypothesis in its original form did lead to certain clearly undesirable structures. For example, affixes excluded from the PW with their root (i.e. "level 2" affixes), as well as clitics, were obligatorily labeled Phonological Words themselves, to permit them to be licensed by the next level, the Clitic Group, as illustrated in the Italian example in (6).

(6)

PW	PW	PW	PW	
[me]	[lo]	[ri]	[sped-isce]	
to me	it	re-	sends	'(he) resends it to me'

In subsequent work, it was argued that the Strict Layer Hypothesis could be slightly weakened such that constituents are no longer required to dominate constituents of the immediately lower level. Thus, the clitics and prefix in a structure such as (6) need not constitute Phonological Words in order to be licensed, effectively removing the main challenge to the CG. Nevertheless, the Clitic Group is often excluded from the hierarchy and other types of structures are used instead, as discussed below.

Another argument that was sometimes used to challenge the Clitic Group as a constituent of the prosodic hierarchy is the fact that in some cases it appears to be superfluous, directly mirroring the lower Phonological Word level, as in (7).

(7)

CG	CG	CG	CG	CG
|	|	|	|	|
PW	PW	PW	PW	PW
Large	green	ideas	sleep	fast

While it is true that in (7) each Clitic Group consists of a single Phonological Word, not all sentences of English can be analyzed in a similar way. Instead, the Clitic Group frequently also includes clitics and affixes excluded from the Phonological Word with their root, analogous to the Italian example in (6).[3] Thus, even if there is some degree of overlap between the CG and the PW, this does not constitute

[3] In fact, the original version of the sentence in (7) would include complex CGs for *colorless* and *furiously*, since the suffixes *–less* and *–ly* are "level 2" affixes and thus not part of the PW of the root. They would have thus also been assigned PW status.

evidence that the Clitic Group is not a constituent of the prosodic hierarchy. Furthermore, even if some language exhibited complete overlap between the PW and the CG, it is not clear that we should conclude that one of these constituents is absent from the language. As pointed out above, if a language lacks an overt manifestation of a particular structure that is considered a component of Universal Grammar, it can still be assumed that the structure is part of a speaker's competence, as opposed to setting up a different class of languages to handle isolated cases.

Crucial to the evaluation of the need for the Clitic Group is the fact that it may serve as the domain of phonological phenomena that are distinct from those that apply in the Phonological Word. For example, in northern varieties of Italian, intervocalic /s/ is voiced within a Phonological Word but not elsewhere. This rule is given in (8); the examples in (9) and (10) show its application (within and across morphemes) and its absence of application (preceding and following the host in a Clitic Group), respectively. The relevant segments are in bold; "z" indicates that the rule has applied and "s" that it has not applied.

(8) Intervocalic s-Voicing: /s/ → [z] / [. . .V _ V . . .]$_{PW}$

(9) Application of Intervocalic s-Voicing

a. [i**z**ola]$_{PW}$ 'island' b. [cau**z**-ano]$_{PW}$ '(they) cause'
c. [fam-o**z**-o]$_{PW}$ 'famous'

(10) Absence of Intervocalic s-Voicing

a. [ri [**s**ala]$_{PW}$]$_{CG}$ '(he) re-salts'
b. [lo **si** [**s**ala]$_{PW}$]$_{CG}$ '(he) salts it for himself'
c. [[guardando]$_{PW}$ **si**]$_{CG}$ 'looking at himself'
d. [[porta]$_{PW}$[**s**apone]$_{PW}$]$_{CG}$ 'soap dish'

As can be seen in (10d), Intervocalic s-Voicing does not apply across the words of a compound, as Phonological Word rules typically do not apply across such a juncture (cf. among many others N&V). It should be noted that despite their PW status, the members of compounds are considered here to constitute a Clitic Group rather than the higher Phonological Phrase constituent. Although this is not observed in the Italian examples, it has been shown elsewhere that such a grouping is necessary since compounds may have different phonological properties from those observed in the PPh as well as the PW (e.g. Vogel and Raimy 2002 for English; Kabak and Vogel 2001 for Turkish). Indeed, it is for this reason that it has been suggested that a more appropriate name for the CG constituent might be "Composite Group", since it includes structures with compounds as well as with clitics and affixes (cf. Vogel 2008).

While Intervocalic s-Voicing applies in the Phonological Word but not the Clitic Group, other rules are observed in the CG but not the PW. For example, in sequences of clitics in Italian, the /i/ of a dative pronominal clitic changes to [e] when followed by a direct object clitic, as in (11). It should be noted that the change is not due to a constraint on the segmental sequences themselves, as seen by the existence of words such as *mila* 'thousands'.

(11) Clitic Vowel Rule

a. m**i** racconta la storia	'(he) tells me the story'
b. m**i** la racconta → m**e** la racconta	'(he) tells it to me'
c. raccontando m**i** la → raccontando m**e** la	'telling it to me'

A straightforward interpretation of the Clitic Vowel Rule is to consider it a Clitic Group domain rule. All that is needed is the additional specification of the particular requirements for its application, as stated in (12).[4]

(12) Clitic Vowel Rule: /i/ → [e] / [. . . [C__]$_{dat}$ [CL]$_{acc}$. . .]$_{CG}$

It is interesting to note that both the Clitic Group and the Phonological Word, the two prosodic constituents that involve an interface with morphology, exhibit idiosyncratic phenomena.[5] We can thus formulate an additional universal:

(13) **Universal 3: Idiosyncratic phonological phenomena are found in the constituents involving an interface with morphology (i.e. the Phonological Word and Clitic Group).**

Similar idiosyncratic phonological phenomena do not appear to be characteristic of the larger constituents. This seems to follow from the fact that such constituents involve an interface with syntax, which itself does not exhibit idiosyncrasies like those of morphology.

3.4 Prosodic Constituents and Optimality Theory

While the approach taken in this paper is essentially derivational, many of the points could also be made in terms of a constraint based model of phonology. Indeed, a number of researchers have made crucial reference to prosodic constituents in OT analyses (cf. among many others Truckenbrodt 1999; Selkirk, 1995, 2000; Anderson 2005). It should be noted, however, that OT analyses do not necessarily involve prosodic constituents per se, since in some cases only a left or right constituent boundary is referred to. That is, there is no positive definition of what is included in a given constituent – only the location of its edge. The important point is, however, that despite such differences, the fundamental insight is similar. That is, it is necessary to recognize a set of phonological phenomena operating in relation to structures of other components of grammar, although the domains for such phenomena are not necessarily isomorphic to the other structures.

---

[4] This rule is formulated somewhat informally here since its formulation is not relevant to the present discussion.

[5] In the PW, these are the "level 1" or "+ boundary" type phonological phenomena in earlier frameworks, although these are not discussed here.

4 Universals: Geometry of Prosodic Constituents

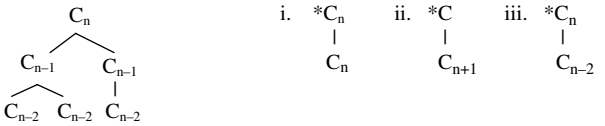
A central aspect of the geometry of the prosodic hierarchy, the Strict Layer Hypothesis, was an early target of research investigating the potential universality of prosodic phonology. It is shown below how relatively small modifications are able to eliminate some early problems with the Strict Layer Hypothesis, while retaining its major insights. Another aspect of the geometry of the prosodic hierarchy that was originally proposed as a type of universal in N&V, the structure of the Phonological Phrase, has been less extensively examined from this perspective. It will be made explicit below how this constituent also contributes interesting insights with respect to universals of the geometry of the prosodic hierarchy.

4.1 The Strict Layer Hypothesis

The Strict Layer Hypothesis is not a single principle, but in fact comprises several sub-components. It is the component referred to as Strict Dominance (cf. (14)) that has led to the most serious challenges to the original geometry of the prosodic hierarchy. According to (14), the structures in (15a) are well-formed, but those in (15b) are not.

(14) Strict Layer Hypothesis: Strict Dominance
A Constituent C_n can only dominate Constituents of level C_{n-1} in the Prosodic Hierarchy.

(15) a. *Well-formed Structure* b. *Ill-formed Structures*

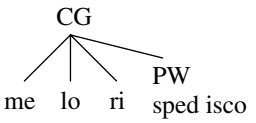


It should be noted that both (15bi) and (15bii) involve recursive structures. In the former, a constituent dominates another constituent of its own level, while in the latter, a constituent dominates a constituent of a level higher than itself. The former, but not the latter, has received a good deal of attention, and it has been proposed by a number of researchers that such recursive structures must be permitted in the prosodic hierarchy (cf. among others, Itô and Mester 1992; Selkirk 1995; Peperkamp 1997; Booij 1996, 1999; Ladd 1996; Truckenbrodt 1999; Anderson 2005). The structure in (15biii) permits a constituent to dominate material more than one level lower in the hierarchy. This structure, too, has been viewed as a possible addition to the prosodic hierarchy by the same researchers as well as others (e.g. Vogel 1999, 2008; Kabak and Vogel 2001).

4.1.1 Skipping Levels in the Prosodic Hierarchy

In (6) above, it was seen that PW status was assigned to such elements as clitics and affixes in accordance with the Strict Layer Hypothesis. These elements, however, typically do not have the same properties as Phonological Words comprising lexical items. For example, they tend not to meet the minimality requirement observed in many languages according to which a Phonological Word must consist of at least two moras. Thus, it was proposed that such items should not, in fact, be considered PWs (cf. among others, Itô and Mester 1992; Vogel 1994, 1999). If this change is made, the consequence is that the next constituent in the prosodic hierarchy must be permitted to dominate material more than one level lower. Thus, (6) would now have a structure such as (16).

(16)



'(he) resends it to me'

This minimal change allows us to retain the other positive aspects of the Strict Layer Hypothesis related to requirements that constituents be properly licensed in specific ways. We can thus formulate the following Universal that reflects this change:

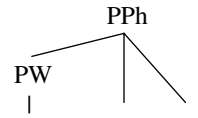
(17) **Universal 4: Prosodic Constituents may dominate items more than one level lower in the hierarchy.**

It remains to be determined whether such a statement can be refined so as to restrict the additional possibilities it introduces. For example, it might be possible to limit the number of levels that may be skipped. It might also be possible that only the Clitic (Composite) Group permits such skipping (cf. Vogel 1999).

4.1.2 Recursivity in the Prosodic Hierarchy – Phonological Word

As mentioned, a number of researchers proposed modifying the Strict Layer Hypothesis to permit recursive structures of the type seen in (15bi). This possibility was exploited by Peperkamp (1997) as a means of accounting for different stress behaviors in several dialects of Italian. While in Standard Italian stress is observed only within the Phonological Word, in the Lucanian and Neapolitan stress may be adjusted in the presence of enclitics. The difference among these three varieties of Italian is represented by Peperkamp in terms of their PW structures, as illustrated below.

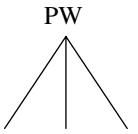
(18) Standard Italian – Recursive PW Analysis (Peperkamp 1997)



host	(CL)	(CL)	
v**é**ndi			‘sell’
v**é**ndi	lo		‘sell it’
v**é**ndi	me	lo	‘sell me it’

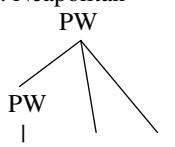
(19) Italian Dialects – Recursive PW Analysis (Peperkamp 1997)

b. Lucanian



host	(CL)	(CL)	
v**í**nnə			‘sell’
vənn**í**	llə		‘sell it’
vinnə	m**í**	llə	‘sell me it’

c. Neapolitan



host	(CL)	(CL)	
c**ó**ntə			‘tell’
c**ó**ntə	lə		‘tell it’
c**ó**ntə	t**í**	llə	‘tell yourself it’

Clitics do not affect the position of word stress in Standard Italian, so they are excluded from the Phonological Word, and joined directly into the Phonological Phrase. Since Peperkamp does not include the Clitic Group in her analysis, at least one level of structure is skipped to permit the direct inclusion of clitics in the Phonological Phrase. A different pattern is observed in Lucanian where the (final) enclitic draws stress to the preceding syllable, whether this is part of the host or another enclitic. According to Peperkamp, this means that enclitics are part of the PW with the host. Neapolitan, by contrast, only shows a stress change if two clitics are present. If one clitic is present, no change takes place, but if two are present, the first becomes stressed; the Phonological Word appears to retain its own stress as well. In this case, it is argued that a recursive Phonological Word structure is needed, where a PW may dominate another PW. It should be noted that this is done specifically in order to account for a distinct phonological behavior that is not observed in the (basic) Phonological Word or the Phonological Phrase, and there is no constituent available for this in a hierarchy lacking the Clitic Group, a point we return to below.

The introduction of the recursive Phonological Word, however, has several drawbacks as can be seen with respect to the Italian examples. First, it leads to a contradiction in the definition of constituent in linguistic theory. A constituent is understood to be a particular type of string, and all constituents of a given type exhibit the same properties, regardless of the size or internal structure of the constituent. If there are two (or potentially more) Phonological Words – an inner PW and an outer PW (or PW’), and these structures have different properties, we are no longer

able to maintain the concept of Phonological Word as a constituent identified by a specific set of properties. In addition, by using the same term, Phonological Word, for both types of structures, we obscure the fact that there are, in fact, different types of strings that exhibit distinct properties.

In the case of Italian in particular, the use of three different structures for stress assignment fails to account for the fact that Standard Italian, as well as Lucanian, Neapolitan and numerous other dialects, all actually share the same generalization regarding the position of primary word stress. Specifically, they all exhibit the "three syllable window", the fact that primary stress falls on one of the last three syllables of a word, and the stress assignment principles responsible for word stress are distinct from those for structures involving clitics (cf. among others Repetti to appear). That is, in words, stress is not always penultimate as in Lucanian structures with any number of clitics, or in Neapolitan structures with two clitics. Furthermore, it is not regularly found more than three syllables from the end word as it is in Neapolitan when a single affix is added to verb forms with antepenultimate stress (e.g. *péttinalə* 'comb them') (Peperkamp, p. 180).[6] It should also be noted that in Neapolitan when two clitics are present, it appears that there are two primary stresses, a possibility that is not normally encountered in Phonological Words, which are considered to exhibit a single primary stress as one of their defining properties.

Assuming the definition of constituent as a linguistic structure with a clearly identifiable property or set of properties, we see that the Phonological Word is identifiable across Italian dialects as a string that exhibits the "three syllable window" for stress. Furthermore, this stress is assigned by a combination of phonological and morphological properties of a word, not simply by counting from the right edge of a word. By contrast, the combination of PW + clitics exhibits different properties, depending on the dialect. If specific phonological properties allow us to identify phonological constituents, and we have two sets of properties, it follows that we should have two constituents. Indeed, Peperkamp and others who make use of recursive Phonological Words, must distinguish between phenomena that apply to the "inner" or "lower" PW and those that apply to a larger "outer" or "higher" PW. Thus, what may seem like single constituent (i.e. Phonological Word) is, in fact, two distinct constituents with a partially overlapping name. In effect, the qualification of type or level of Phonological Word is merely a diacritic to account for the observed differences in behavior.

An alternative analysis is available, however, that makes use only of a single innovation with respect to the Strict Layer Hypothesis, the possibility of skipping levels in the prosodic hierarchy. If we assume the presence of the Clitic (Composite) Group, it is possible to include within this constituent a Phonological Word plus any excluded material (i.e. certain affixes and clitics), as in (20)–(21), where brackets are used in place of trees.

---

[6] Stress also appears (atypically) in Standard Italian on the pre-antepenultimate syllable of a small set of verbs, although such forms are rare and independent of the presence of clitics (e.g. *péttinano* '(they) comb', *péttinino* '(that they) comb'). Given their specific and limited distribution, such forms are not taken as evidence against the "three syllable window".

(20) Standard Italian – Composite Group Analysis

[[host]$_{PW}$ (CL) (CL)]$_{CG}$	
v**é**ndi	‘sell’
v**é**ndi lo	‘sell it’
v**é**ndi me lo	‘sell me it’

(21) Italian Dialects – Composite Group Analysis

a. Lucanian			b. Neapolitan			
[[host]$_{PW}$ (CL) (CL)]$_{CG}$			[[host]$_{PW}$ (CL) (CL)]$_{CG}$			
v**í**nnə	‘sell’		c**ó**ntə		‘tell’	
vənn**í** llə	‘sell it’		c**ó**ntə	lə	‘tell it’	
vinnə m**í** llə	‘sell me it’		c**ó**ntə	t**í** llə	‘tell yourself it’	

Aside from straightforwardly revealing distinct domains for the different stress properties of Phonological Words and structures with clitics, the proposed analysis has the further advantage of allowing us to account for the stress properties of Standard Italian and the dialects in the same way. That is, in each case, the PW can be identified as the domain of the “three syllable window”. Furthermore, the Composite Group has the same structure in each case, clearly revealing that the difference among the varieties of Italian lies solely in the formulation of the CG domain stress rules in (22)–(23). Since the CG in Standard Italian does not appear to have any distinct stress properties, it may not have a CG level stress rule. Nevertheless, (22) provides a potential rule which assigns CG stress to the primary stressed syllable of the PW regardless of the presence or absence of clitics. This permits a more direct comparison of the facts across dialects.

(22) Standard Italian – CG Stress Rule: $\sigma \rightarrow [+\text{stress}] \,/\, \_\_ \ldots \,]_{PW} \ldots \,]_{CG}$
$[+\text{str}]$

(23) Italian Dialects – CG Stress Rules
a. Lucanian: $\sigma \rightarrow [+\text{stress}] \,/\, \_\_\ \sigma]_{CG}$
b. Neapolitan: $\sigma \rightarrow [+\text{stress}] \,/\,]_{PW}\ \_\_\ \sigma]_{CG}$

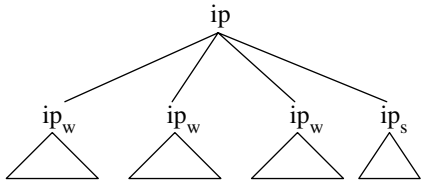
In sum, we have seen that if the Strict Layer Hypothesis is revised slightly to permit the skipping of levels in the prosodic hierarchy, it is not necessary to also permit recursivity, at least at the PW level, assuming the CG as a constituent in the hierarchy. As shown, there are several advantages to such an analysis including the fact that it clearly identifies distinct constituents when distinct phonological phenomena are observed. In addition, it permits a similar characterization of the domains involving an interface with morphology as precisely those two domains that allow phonologically idiosyncratic phenomena, reflecting the fact that idiosyncrasy is considered a property of the morphological component of grammar.

4.1.3 Recursivity in the Prosodic Hierarchy – Intonational Phrase

Another constituent in the prosodic hierarchy that has been identified as permitting recursivity is the Intonational Phrase, most notably in the work of Ladd. Thus,

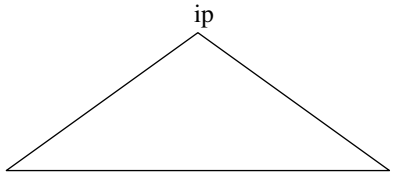
structures such as (24) are proposed, where the Intermediate Phrase (ip) is permitted to dominate other ips (Ladd 1996: 251). This is crucially distinct from the structure in (25), as indicated by the different meanings associated with the two structures. In the first, "butcher" can be interpreted as an epithet for an incompetent surgeon, while in the second it refers literally to the person selling meat.[7]

(24)



The butcher charged me a thousand bucks.

(25)



The butcher charged me a thousand bucks.

It should be noted that Ladd's ips, and to some extent N&V's Intonational Phrases, are different from the other prosodic constituents in that they delimit domains over which intonational contours are spread. In fact, Ladd (1996: 246) proposes that the ip may be defined straightforwardly as "nothing more than the smallest prosodic unit that can have a tune." By contrast, the other constituents, as well as N&V's Intonational Phrase, delimit domains of phenomena that target specific segments or syllables (e.g. segmental rules, phonotactic constraints, stress assignment, tone rules). The IPh/ip also exhibits substantially more variability in its formation than the other constituents. This is most likely a reflection of the fact that its construction involves more variable aspects of language such as speech rate, constituent length, focus and other semantic and pragmatic aspects of a sentence. This contrasts with the other prosodic constituents which are constructed in relation to smaller, fixed, sets of morphological and syntactic properties. Indeed, Ladd argues that it is not desirable to try to associate ip structure with particular morphological and syntactic structures. Instead, we should be able to identify ip boundaries on the basis of purely phonetic and

[7] Ladd's ip is defined differently from the Intonational Phrase in N&V, however, it accounts for several of the same phenomena, and the present discussion holds in most cases for both the Intonational Phrase and the ip.

phonological information, "without as it were looking over our shoulder at the theoretical consequences for prosodic structure or for syntax-prosody mapping" (p. 246).

In N&V it was originally proposed that the Intonational Phrase serves as the domain of phonological phenomena of the types observed in the other constituents, as well as serving as the domain for intonational contours. Given more recent insights regarding the associations between IPhs/ips and intonational contours, it might turn out to be the case that the Intonational Phrase is not actually the domain of specific phonological rules and constraints. Instead, the cases in N&V where the Intonational Phrase was claimed to be the domain for segmental rules may be found to reflect the fact that the ends of the strings in question exhibit boundary tones for semantic/pragmatic reasons, and these in turn lead to the absence of application of a particular phonological rule across such a juncture. Such a possibility, however, remains to be examined in more detail.

Rather than introduce recursivity into the prosodic hierarchy, however, an alternative analysis of such phenomena is possible. This one, too, requires a departure from the original set and arrangement of prosodic constituents, although, it does not require changes in the basic geometry of the prosodic hierarchy. Given the substantially different nature of the Intonational Phrase with respect to the other prosodic constituents, it seems that the IPh may not, in fact, be part of the prosodic hierarchy as originally proposed. That is, in addition to involving more flexible factors that lead to significantly more variability than seen in other types of constituents, the IPh/ip appears to interact directly with the other prosodic constituents, as opposed to representing a domain that dominates smaller constituents or is dominated by larger constituents. As a constituent based on purely phonological and phonetic considerations, as Ladd proposes, it is possible, furthermore, that the IPh/ip more appropriately belongs in a hierarchy with other purely phonetic/phonological constituents including the mora, syllable and foot, although these structures are not themselves recursive.

Regardless of its possible placement in a different hierarchy, if the IPh/ip is no longer a constituent of the prosodic hierarchy with the constituents that involve interfaces with morpho-syntactic structures, there is again no need to introduce recursivity into the geometry of the hierarchy. We can thus formulate the additional Universal given in (26). Consequently, we must also modify Universal 2 to exclude the Intonational Phrase from the set of prosodic constituents that involve an interaction with morphology and syntax.

(26) Universal 5: Prosodic Structures are not recursive.

By removing the Intonational Phrase, as well as the associated recursivity, from the original prosodic hierarchy, we are able to retain the more restrictive nature of the phonological interface with morphology and syntax found in the other constituents. Furthermore, the removal of recursivity at the Intonational Phrase level as well as at the Phonological Phrase level (e.g. Truckenbrodt 1999), discussion of which is beyond the scope of the present paper, brings the prosodic hierarchy in line with recent claims that while recursivity is a crucial aspect of syntax, it is not an aspect of phonology (e.g. Jackendoff and Pinker 2005; Pinker and Jackendoff 2005;

Neeleman and van de Koot 2006). It is possible that the fact that intonational structures appear to be recursive reflects their dependence on semantics and pragmatics, which need not be restricted in the same way as phonological structures.

4.2 Phonological Phrase Structure

The Phonological Phrase is the constituent in the prosodic hierarchy that involves the closest interaction with syntax.[8] In fact, it has been proposed that the Phonological Phrase plays a crucial role in language acquisition, permitting "bootstrapping" to the syntactic structure of a language (cf. among others, Nespor et al. 1996). The original definition of the Phonological Phrase formulated in terms of X' syntax is that shown in (27), with minor modifications in wording.

(27) Phonological Phrase Construction (N&V, p. 168)
The domain of the Phonological Phrase consists of a Clitic Group (CG) which contains a lexical head (X) and all CGs on its non-recursive side up the next CG that contains another head outside of the maximal projection of X.

Languages may vary with respect to the inclusion of complements in the Phonological Phrase. As seen above in (2), French excludes complements from the Phonological Phrase. Thus, liaison applies between an adjective and a following noun (un savan[t] anglais 'a wise Englishman') but not between a noun and a following adjective (un savan[Ø] anglais 'an English scholar'). Clements (1978) reports that Ewe also excludes complements on the recursive side, as evidenced by the application of several tone rules.

Italian differs from French and Ewe in allowing a complement on the recursive side of a head to be included in the Phonological Phrase as long as it is non-branching. Thus, Raddoppiamento Sintattico (RS), the phenomenon that essentially geminates a word initial consonant following a stressed vowel is observed in (28a) between an adjective and a noun, where there is no complement on the recursive side. In (28b), we observe RS between a noun and a following single adjective, a non-branching complement on the recursive side. RS does not apply, however, in (28c), where the complement on the recursive side is branching. The acute accents indicate the final stressed vowels in question; the relevant segments are in bold.

(28) Raddoppiamento Sintattico

a. [i tré ggatti]PPh	'the three cats'
b. [il caffé fforte]PPh	'the strong coffee'
c. [il caffé]PPh [poco forte]PPh	'the not-so-strong coffee' (*poco* = 'little')

[8] The Phonological Phrase has led to the positing of the most related but slightly different constituents, including major phrases, focus phrases, intermediate phrases, and quantity phrases. In many cases, the criteria for defining such constituents, and the function of the constituents, are somewhat different from those in N&V. For the present purposes, the discussion will be limited to the characteristics of the PPh as proposed in N&V.

A different possibility is observed in Chimwi:ni, a Bantu language, where the first complement on the recursive side of a head may be included in the Phonological Phrase regardless of its length (cf. Hayes 1989). Odden (1990) reports a similar pattern in another Bantu language as well, Kimatuumbi.

Given the range of possibilities in the mapping from syntax onto Phonological Phrase structure, we may first formulate a general universal with regard to the inclusion of complements on the non-recursive side of a head, based on N&V, as in (29). Additionally, it seems possible that there is an implicational relationship among the options available for the construction of the PPh, as indicated in (30).

(29) **Universal 6 (N&V): If a language includes complements in the Phonological Phrase, these will be on the non-recursive side of the head.**

(30) **Universal 7: Inclusion of Complements**
If a language permits non-branching complements in the Phonological Phrase, it will also have Phonological Phrases with no complements; if a language permits branching complements in the Phonological Phrase, it will also have Phonological Phrases with non-branching complements (and no complements).

The universal in (30) is analogous to another, well-known, implicational universal that applies to syllable structure. Specifically, the first part is similar to the implication *CVC* > *CV*, that is, if a language has CVC syllables, it will also have CV syllables. The second part is similar to the additional implication C*VCC* > *CVC* > *CV* (cf. Greenberg et al. 1963). Even if such statements only represent tendencies, it is interesting that the same types of implicational relationships are observed in different aspects of the phonological system.

5 Universals: Prosodic Structure Phenomena

The constituents defined within the theory of prosodic phonology delimit the domains of application of any type of phonological phenomenon that involves information beyond the properties of individual segments. Thus, we find a wide variety of phenomena analyzed within the prosodic phonology model such as a) segmental rules (e.g. English Flapping), b) phonotactic constraints (e.g. Italian Phonological Words may not begin with the palatal lateral [λ]), c) stress assignment (e.g. in numerous languages each Phonological Word has one primary stress), d) stress adjustment (e.g. English Rhythm Rule), and tone rules (e.g. Chinese Tone Sandhi, African tone rules). The question that arises at this point in how such phenomena may be formulated.

5.1 Prosodic Domain Rules

As proposed in works such as Selkirk (1980, 1981, 1986) and N&V, prosodic phonological phenomena will exhibit one of three rule formats, leading to the positing of the next universal. In (31), Di and Dj refer to prosodic constituents, where Dj is higher in the hierarchy than Di.

(31) **Universal 8: Prosodic Phenomena will have one of the following formats.**
a. Domain Edge: $[\ldots\_\_]_{Di}$
b. Domain Span: $[\ldots\_\_\_\ldots]_{Di}$
c. Domain Juncture: $[\ldots[\ldots\_\_]_{Di}[\ldots]_{Di}\ldots]_{Dj}$

These rule formats were essentially assumed to be one of the core components of prosodic phonology, and thus implicitly considered universals. Closer examination shows that they in fact provide a rich set of possibilities for the application of phonological phenomena, suggesting additional possibilities for universals related to the application of prosodic phenomena. One such possibility involves another implicational universal, as seen in (32), and its corollary in (33).

(32) **Universal 9: If a language has Domain Juncture Rules, it will also have both Domain Span and Domain Edge Rules.**
(33) **Corollary: The presence of Domain Edge and Domain Span Rules does not necessarily imply the presence of corresponding Domain Juncture Rules.**

In fact, the predictions made by (32) and (33) are not surprising if we consider the nature of Domain Juncture Rules. That is, they involve precisely the adjacency of two domain edges, and this configuration must be confined within some larger string or domain span. This type of implication is analogous to the well-known implicational universal regarding complex segments and its corollary proposed by Jakobson (cited by Hockett 1963). That is, it appears that if a language has an affricate at a given place of articulation, it also has corresponding stops and fricatives. The corollary is that the presence of stops and fricatives at a given place of articulation does not necessarily imply the presence of corresponding affricates.

5.2 Hierarchical Application of Prosodic Phenomena

Another possible universal with regard to the application of prosodic phenomena directly appeals to the hierarchical nature of the model, although such claims were rarely made in the original proposals (but cf. Vogel 1984; Rice 1990). That is, since smaller constituents are embedded within larger ones, it seems reasonable that the relative size or place of a constituent in the hierarchy will result in a relation between the domains of application of certain phenomena. Thus, we can posit the following universal, which is illustrated below in relation to phonotactic constraints of English and Italian.

(34) **Universal 10: If a constraint holds for a Prosodic Domain D, it also holds for smaller domains.**

In English, the velar nasal [ŋ] is excluded from the onset position of (phonological) words, as indicated in (35).

(35) English Phonotactic Constraint: *[ŋ] / [__ . . .]$_{PW}$

Thus, while words and larger structures begin with [m] and [n], the velar nasal is absent not only at the beginning or words, but also larger constituents, as shown in (36), where the relevant segments are in bold:

(36) a. [$_{CG}$ a [$_{PW}$ **m**at]] but *[$_{CG}$ a [$_{PW}$ **ŋ**at]]

b. [$_{PU}$[$_{PPh}$[$_{CG}$[$_{PW}$ **n**umerous]] [$_{CG}$[$_{PW}$ **m**ats]]] [$_{PPh}$[$_{CG}$[$_{PW}$ **m**asked]] [$_{CG}$ the [$_{PW}$ **m**ess]]]]

It might appear possible to extend the restriction on [ŋ] to apply to syllable onsets within the PW, however, it is not clear that the same phonological phenomena apply to both the constituents identified earlier as "purely phonological" (e.g. syllable) and to those under consideration here involving interfaces with other components of grammar. For example, in a word like *dinghy* [dIŋ i], if we syllabify the segments as CVCV, [ŋ] would be in the onset position of a syllable. It might be possible to argue that in such cases [ŋ] becomes ambisyllabic and thus is also the coda of the previous syllable, but what remains clear is that there is a difference in the applicability of a constraint such as (35) to syllables within words and those that are onsets of the interface constituents, beginning with the PW.

One question that arises is whether the universal in (34) should be formulated such that constraints holding in larger constituents also hold in smaller ones, or vice versa, that is, that constraints holding in smaller constituents also hold in larger ones. With regard to the English constraint, in fact, it would appear just as effective to posit a constraint at the highest level of the hierarchy, the Phonological Utterance (PU): *[ŋ]/[__. . .]$_{PU}$. We could then formulate the universal, instead, to apply also to all smaller (interface) constituents, down to the Phonological Word.

While the English case does not permit us to choose between the formulation of Universal 10 as given in (34) and its alternative formulation, a phonotactic constraint of Italian does permit us to choose. In Italian, the palatal lateral [λ], orthographically "gl", is found within words (e.g. *figlio* 'son'), but not at the beginning of lexical items. It does, however, appear at the beginning of clitics (e.g. *gli* 'to/for him').[9] In terms of prosodic constituents, this amounts to the exclusion of [λ] at the beginning of Phonological Words, but not other constituents, as expressed in the following phonotactic constraint:

(37) Italian Phonotactic Constraint: *[λ]/[__. . .]$_{PW}$

[9] The absence of [λ] in final position is simply due to the avoidance of final consonants in general in Italian.

This correctly predicts that while we observe a restriction on [λ] at the onset of the Phonological Word constituent, and potentially any smaller (interface) constituents if they existed, this restriction is not observed at the onset of larger constituents, as illustrated in (38).

(38) a. [$_{PW}$ **l**eggo] '(I) read' but *[$_{PW}$ [λ]eggo]
b. [$_{CG}$ **gl**i [$_{PW}$ correggo]] '(I) correct for him'
c. [$_{PU}$ [$_{PPh}$ [$_{CG}$ **gl**i [$_{PW}$ correggo]] [$_{CG}$ [$_{PW}$ sempre]]][$_{PPh}$ [$_{CG}$ **gl**i [$_{PW}$ sbagli]]]]
'(I) always correct the errors for him.'

As can be seen, the palatal lateral appears not only at the beginning of the constituent above the Phonological Word, the Composite (Clitic) Group, but consequently also at the beginning of each of the larger prosodic constituents. It is thus not the case that if [λ] is excluded at the level of the Phonological Word, it is also excluded at each higher level, as we were able to state for English.

What is true in both cases, however, is that once we identify the largest domain for a particular phonotactic constraint, the Phonological Utterance for English and the PW for Italian, we can predict that the same constraint holds for each smaller constituent. Again, it should be noted that we are not considering "purely" phonological constituents such as the syllable to be subject to the same phonological phenomena as the interface constituents.[10]

6 Conclusions

The systematic study of prosodic structures is relatively recent, and the study of universals of these structures is even more recent. In this paper, several possible universals have been extracted from the original presentation of phonology in N&V. Additional universals have been proposed based on further examination of the model as well as certain proposed modifications, in particular with relation to the Strict Layer Hypothesis.

The ten universals proposed here fall into three general categories relating to (a) general properties of prosodic structure, (b) the geometry of prosodic constituents, and (c) prosodic structure phenomena. With respect to Greenberg's original types of universals, both simple universals (i.e. all languages have/do not have X) and implicational universals (i.e. if a language has X, it also has Y) have been proposed. It was pointed out that some of the aspects of prosodic phonology, and thus the related universals posited here, have been challenged and in some cases have led to modifications of the theory. It is important in each case to consider whether apparent

[10] Flack (2007) addresses in detail the relationships among phonotactic constraints that apply at different levels of the prosodic hierarchy within the framework of Optimality Theory, and includes more possibilities than those predicted by Universal 10. While this might indicate counter-evidence to Universal 10, it should be noted that Flack does not use precisely the same prosodic structures (e.g. she excludes the Clitic Group). As a result, at least some of the differences may be due to factors other than the issue raised by Universal 10, a matter that requires further investigation.

violations of universals are in fact true violations or whether they may be due to inadequate data, a problem associated in particular with the larger constituents of the prosodic hierarchy involving interfaces with syntax and possibly semantics. If significant deviations are confirmed, it is still possible that the proposed universals represent general tendencies observed widely across languages as opposed to absolute universals. It remains now to further test the proposed universals, as well as to determine if there are additional properties of prosodic structures that qualify as universals of prosodic structure.

References

Anderson, Stephen. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

Basbøll, Hans 197. Grammatical boundaries in phonology. *Aripuc* 9: 109–135.

Basbøll, Hans. 1981. On the function of boundaries in phonological rules. In *Phonology in the 1980's*, ed. Didier Goyvaerts, 245–269. Ghent: Story-Scientia.

Bickel, Balthasar, and Kristine Hildebrandt. 2007. Word domains. Cross- linguistic challenges for the prosodic hierarchy: evidence from word domains. Ms. University of Leipzig.

Booij, Geert. 1996. Cliticization as prosodic integration: the case of Dutch. *The Linguistic Review* 13: 219–242.

Booij, Geert. 1999. The role of the prosodic word in phonotactic generalizations. In *Studies on the Phonological Word*, ed. Tracy A. and Ursula Kleinhenz, 47–72 Philadelphia: John Benjamins.

Booij, Geert. 2007 [2nd edition]. *The Grammar of Words*. Oxford: Oxford University Press.

Carter, Richard Jr. 1974. *Teton Dakota Phonology*. PhD. Dissertation. University of New Mexico. Published as *University of Manitoba Anthropology Papers*. 10.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Clements, George N. 1978. Tone and syntax in Ewe. In *Elements of Tone, Stress and Intonation*, ed. Donna Jo Napoli. Washington, DC: Georgetown University Press.

Comrie, Bernard. 1981. *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.

Dixon, Robert M. W., and Alexandra Y. Aikhenvald. 2002. Word: a typological framework. In *Word* eds. Robert M. W. Dixon and Alexandra Y. Aikenvald, 1–41. Cambridge: Cambridge University Press.

Downing, Laura. 1999. Prosodic stem and prosodic word in Bantu. In *Studies on the Phonological Word*, eds. Tracy A. Hall and Ursula Kleinhenz, 73–98. Philadelphia: John Benjamins.

Flack, Katherine. 2007. *The Sources of Phonological Markedness*. PhD. Dissertation. University of Massachusetts.

Greenberg, Joseph, Charles Osgood, and James Jenkins. 1963 [1966 second edition]. Memorandum concerning language universals. In *Universals of Language*, ed. Joseph Greenberg, xv–xxvii. Cambridge, MA: MIT Press.

Hall, Tracy A. 1999. The phonological word: a review. In *Studies on the Phonological Word*, eds. Tracy A. Hall, and Ursula Kleinhenz, 1–22. Philadelphia: John Benjamins.

Hall, Tracy A., and Ursula Kleinhenz (eds.). 1999. *Studies on the Phonological Word*. Philadelphia: John Benjamins.

Hockett, Charles. 1963 [1966 second edition]. The problem of universals in language. In *Universals of Language*, ed. Joseph Greenberg, 1–29. Cambridge, MA: MIT Press.

Hayes, Bruce. 1989 [1984]. The prosodic hierarchy in meter. In *Rhythm and Meter*, eds. Paul Kiparsky, and Gilbert Youmans, 201–260. Orlando, Florida: Academic Press.

Itô, Junko, and Armin Mester. 1992. Weak Layering and Word Binarity. Ms. University of California at Santa Cruz.

Jackendoff, Ray, and Stephen Pinker. 2005. The nature of the language faculty and its implications for the evolution of language. (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97: 211–225.

Kabak, Baris, and Irene Vogel. 2001. Stress in Turkish. *Phonology* 18: 315–360.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Neeleman, Ad, and Hans van de Koot. 2006. On syntactic and phonological representations. *Lingua* 116: 1524–1552.

Nespor, Marina, Teresa Guasti, and Anne Christophe. 1996. Selecting word order: the Rhythmic Activation Principle. In *Interfaces in Phonology*, ed. Ursula Kleinhenz, 1–26. Berlin: Akademie Verlag.

Nespor, Marina, and Irene Vogel. 1982. Prosodic domains of external Sandhi rules. In *The Structure of Phonological Representations*, vol. 1, eds. Harry van der Hulst, and Norval Smith, 225–255. Dordrecht: Foris.

Nespor, Marina, and Irene Vogel. 1983. Prosodic structure above the word. In *Prosody: Models and Measurements*, eds. Anne Cutler, and D. Robert Ladd, 123–140. Heidelberg: Springer.

Nespor, Marina and Irene Vogel. 1986. *Prosodic Phonology*. Dordrecht: Foris Publications.

Odden, David. 1990. Syntax, lexical rules and postlexical rules in Kimatuumbi. In *The Phonology-Syntax Connection*, eds. Sharon Inkelas, and Draga Zec, 259–277. Chicago: University of Chicago Press.

Peperkamp, Sharon. 1997. *Prosodic Words*. The Hague: Holland Academic Graphics.

Pinker, Stephen and Ray Jackendoff. 2005. The faculty of language: What's Special about it? *Cognition* 95: 201–236.

Repetti, Lori. To appear. *The Prosodic Structures of Northern Italian Dialects*. Amsterdam: Kluwer.

Rice, Keren. 1990. Predicting rule domains in phrasal phonology. In *The Phonology-Syntax Connection*, eds. Sharon Inkelas, and Draga Zec, 289–312. Chicago: University of Chicago Press.

Schiering, René, Kristine Hildebrandt, and Balthesar Bickel. 2007. Cross-linguistic challenges for the prosodic hierarchy: evidence from word domains. Ms. University of Leipzig.

Selkirk, Elizabeth. 1972. T*he Phrase Phonology of English and French*. Ph.D. Dissertation: MIT Press. (Published in 1980 by Garland Press)

Selkirk, Elisabeth. 1980. Prosodic Domains in Phonology: Sanskrit Revisited. In *Juncture*, eds. Mark Aronoff, and Mary-Louise Kean, 107–129. Saratoga: Anma Libri.

Selkirk, Elizabeth. 1981 [1978]. On prosodic structure and its relation to syntactic structure. In *Nordic Prosody II*, ed. Thorsten Fretheim, 111–140. Trondheim: TAPIR.

Selkirk, Elizabeth. 1986. On derived domains in sentence phonology. *Phonology Yearbook* 3: 371–405.

Selkirk, Elizabeth. 1995. The prosodic structure of function words. In *Papers in Optimality Theory*, eds. Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 349–370. University of Massachusetts Occasional Papers in Linguistics. Amherst, MA: Linguistics Department.

Selkirk, Elizabeth. 2000. The interaction of constraints on prosodic phrasing. In *Prosody: Theory and Experiments*, ed. Merle Horne, 231–261. Dordrecht: Kluwer.

Simon, Horst, and Heike Wiese (eds.). To appear. *Expecting the Unexpected: Exceptions in Grammar*. Trends in Linguistics: Studies and Monographs Series. Amsterdam: Mouton de Gruyter.

Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry* 30: 219–255.

Vogel, Irene. 1984. On constraining prosodic rules. In *Advances in Nonlinear Phonology*, eds. Harry van der Hulst, and Norval Smith, 217–233. Dordrecht: Foris Publications.

Vogel, Irene. 1994. Phonological Interfaces in Italian. In *Issues and Theory in Romance Linguistics: Selected Papers from the Linguistic Symposium on Romance Languages XXIII*, ed. Michael Mazzola, 109–125. Washington, DC: Georgetown University Press.

Vogel, Irene. 1999. Subminimal constituents in prosodic phonology. In *Issues in phonological structure*, eds. Stephen J. Hannahs and Michael Davenport, 249–267. Amsterdam: John Benjamins.

Vogel, Irene and Eric Raimy. 2002. The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language* 29: 225–250.

Vogel, Irene. 2008. The Morphology-Phonology Interface: Isolating to Polysynthetic Languages". In *Selected Papers from the 12*th *International Morphology Meeting* [Special Issue of *Acta Linguistica Hungarica* 54:2], eds. Ferenc Kiefer and Péter Siptar. Budapest, Hungary.

Lexical Integrity as a Formal Universal: A Constructionist View

Geert Booij

Abstract This paper deals with an important formal universal with respect to the interface of morphology and syntax, the Lexical Integrity Principle. This principle encompasses both non-interruptability and non-accessibility of word-internal structure. Non-interruptability is a defining property of canonical wordhood, and this part of Lexical Integrity is therefore almost never violated. Non-accessibility of word-internal structure should be rejected on empirical grounds, In a constructionist view of morphology, the possibility of syntax and semantics having access to word-internal structure is to be expected.

Keywords Construction grammar · inflection · lexical integrity · morphology-syntax interface · morphological universals · word formation

1 Introduction: Morphological Universals

Morphology has always played an important role in language typology, that is in the systematic characterisation of variation between languages, and the constraints on that variation. The best known classical form of morphological typology is the ranking of languages by means of two indices, the index of synthesis and the index of fusion (Comrie 2001). In more recent work on word formation, in particular compounding that has been inspired by the Principles-and-Parameters framework of Chomsky (1981), the idea of a head-parameter (a language has either right-headed or left-headed compounds) has shown to be a fruitful typological perspective (cf. Scalise ed., 1992). However, it is hard to find uncontested substantive morphological universals of an absolute nature, certainly not in the domain of word formation. This may be a reflection of the fact that the building blocks of word formation patterns often derive from lexical items, through the process of grammaticalization. Hence, word formation patterns will reflect the language-specificity of the lexicon.

G. Booij
University of Leiden

I would like to thank Alice Harris and the anonymous reviewers for their constructive comments on a previous draft of this paper.

S. Scalise et al. (eds.), *Universals of Language Today*, 83–100,
DOI: 10.1007/978-1-4020-8825-4_5,

The discussion of morphological universals in Greenberg (1963), the publication that forms the historical background for the studies in this book, focuses on two issues. The first is that of affix order. Greenberg showed that derivational morphemes tend to be closer to the root than inflectional morphemes. The tradition of investigating the order of derivation and inflection, and regularities in affix ordering has been continued in the work of Bybee, for instance Bybee (1985) and Bybee et al. (1994).

The second issue broached by Greenberg is that of the morphological marking of inflectional categories. His work focused on the morphological asymmetries in the relation between form and meaning. For instance, Greenberg observed such an asymmetry for the category of number: if in languages with two values for number, singular and plural, only one of the values is formally marked, this is usually the plural. Such generalizations are expressed in the form of implicational universals. An excellent survey of such universals can be found in The Universals Archive (http://typo.uni-konstanz.de/archive) developed by Frans Plank and his co-workers.

Greenberg's generalizations on the order of inflectional categories can be accounted for by making a distinction between contextual and inherent inflection. Contextual inflection is the kind of inflection that is required to be present by syntactic contexts, such as person and number marking on finite verbs, gender and number marking on adjectives, and the marking of structural case on nouns. Inherent inflection is inflection that is determined by semantic considerations, such as the marking of number and semantic case on nouns. Arguments for this distinction are presented in Booij (1994, 1996). The generalization then is that contextual inflection is peripheral to inherent inflection, whereas inherent inflection is peripheral to derivational morphological constituents. For instance, the marking of structural case on nouns is peripheral to that for number, and both are peripheral to derivational morphemes in a complex noun. The following example from Hungarian illustrates the ordering of inherent inflection and contextual inflection: the accusative suffix *–t* (an instance of contextual inflection) is preceded by two suffixes that express categories of inherent inflection: the number and the possessor of the noun:

(1) gyereke-i-nke-t
child-PL-1PL-ACC
'our children'(acc.)

Whereas absolute universals are hard to find in the domain of word formation, there is certainly evidence for implicational universals. Recent work on morphological universals of the implicational type can be found in Haspelmath (2006, 2007). An example of such an implicational universal in the domain of word formation that deals with the formation of causative verbs is the following: "If a language has causative verbs derived from transitive bases, then it also has causatives derived from intransitive bases", a generalization that Haspelmath (2007) ascribes to the Russian linguists Nedjalkov and Sil'nickij.

At a more fundamental level the issue is how such universals can be explained. There are a number of types of explanation that have been invoked: cognitive and pragmatic factors such as Relevance, Iconicity, and Economy (as in the work

of Bybee and the theory of Natural Morphology), processing factors (Cutler and Hawkins 1988), and universal mechanisms and pathways of linguistic change (Anderson 2004, Harris 2004, Haspelmath 2006, Bybee, 2006).

As observed by Helmbrecht (2004: 1248), "cross-linguistic generalizations in contemporary morphology are largely generalizations over form-function relations in morphological units" of the sort proposed by Greenberg, such as the generalization about the formal expression of singular and plural mentioned above. Helmbrecht also notes that "[t]here are practically no substantive, i.e. absolute universals with regard to morphology" (Helmbrecht 2004: 1250). This suggests that we have to look for formal universals, if we want to find morphological universals at all. Universals serve to define the notion 'possible natural language' and hence they may have the form of constraints. We may distinguish the following two (related) types of formal universal constraints in the domain of morphology: (i) constraints on the kind of relations that are possible between syntax and word structure, and (ii) constraints on the accessibility of the internal structure of complex words for modules of the grammar such as the syntax and semantics. Such constraints will be discussed in the next section.

2 Constraints on the Interaction Between Syntax and Word Formation

There are two well-known constraints on the interaction between syntax and word formation in the literature, the No Phrase Constraint and the Lexical Integrity Constraint.

As to the No-Phrase Constraint, a good summary of the discussion of this constraint can be found in Lieber and Scalise (2006). They show that this constraint is incorrect: certain types of phrases can form part of complex words, as has been shown in many publications referred to in their article. Such facts receive a straightforward interpretation in a modular and constraint-based grammar. The morphological module specifies well-formedness constraints on complex words. This may include the occurrence of certain types of phrases such as $[AN]_{NP}$ in the non-head position of complex words, as illustrated by the Dutch compound $[[[oude]_A\ [mannen]_N]_{NP}\ [huis]_N]_N$ 'old men's home'. It is the morphological module that defines which kind of phrases can appear within complex words. The syntactic module in its turn defines the well-formedness of those word-internal phrases. Hence, these two modules have to operate in a parallel fashion. Thus, such facts pose a problem to the view of the grammar as a set of ordered components.

The Lexical Integrity Hypothesis is a constraint on the interface between rules/constraints of the grammar and the internal structure of complex words (for recent discussions see Lieber and Scalise 2006). Anderson is one of the morphologists who proposed that syntax has no access to word-internal morphological structure:

(2) *Principle of Lexical Integrity*
"The syntax neither manipulates nor has access to the internal structure of words" (Anderson 1992: 84)

In recent work on the interface between syntax and morphology by Ackema and Neeleman (2005), essentially the same position is defended. In their model of the grammar, sentence grammar and word grammar are different parts of the grammar that only touch each other at the level of lexical insertion where the features of syntactic nodes have to match those of (simplex or complex words). For example, it is necessary for a proper account of agreement phenomena for syntax to have access to the feature specification of a noun for the morpho-syntactic category Number. However, it is not relevant how this feature is expressed morphologically. For instance, for the purpose of number agreement, it does not matter whether the plural suffix of Dutch nouns is *-s* or *-en*; we only need to know if a noun is singular or plural.

The principle of Lexical Integrity as formulated in (1) excludes two kinds of syntax-morphology interaction: (i) manipulation of parts of word-internal structure, and (ii) access to word-internal structure. Under manipulation I subsume the syntactic movement of a word constituent, and the splitting up of words by intermediate constituents. I consider the impossibility of syntactic movement of the constituents of a linguistic unit as a necessary condition for that linguistic unit to be a word. As to the possibility of splitting up a complex word, we will see below that this is a very rare phenomenon.

We need the prohibition on the movement of word constituents for explaining why in Dutch and German the rule of Verb Second that places finite forms of verbs in second position in root clauses cannot strand the prefix of a complex verb such as *doordénken* 'to think through completely', whereas the particle in particle verbs such as *dóordenken*'to continue thinking' can be stranded:[1]

(3) Jan door-dacht het problem/*Jan dacht het probleem door 'John thought about the problem thoroughly'
(4) Jan dacht door over het probleem 'John continued thinking about the problem'

Thus, this part of the Lexical Integrity principle may serves as a basic test to find out if a sequence of morphemes is a word or a phrasal lexical unit (cf. Bresnan and Mchombo 1995). Particle verbs in Dutch, though clearly lexical units, cannot be words since the particle can be separated from the verb as illustrated above. Hence, we have to conclude that they are not words, but phrasal units. A phrasal analysis of the Dutch particle verbs is argued for in Booij (2002) and Blom (2005).

This also holds for the Hungarian pseudo-verbal compounds such as *tévét nez* 'be engaged in television watching' discussed in Kiefer (1992). The two parts of this lexical unit can be split in certain syntactic contexts, for instance by the negative

[1] The rules of Dutch orthography require the particle verb to be spelled as one word, without internal spacing. This reflects the status of 'lexical unit' of particle verbs, but obscures the fact that a particle verb is not one grammatical word.

word *nem* 'not'. We take the separability of these units as evidence for them not being words. This is confirmed by the fact that the noun constituent *tévét* in this example is marked with accusative case (the suffix *-t*). This assignment of structural case to the word *tévé* shows that it must be an independent word. Given the principle of Lexical Integrity, one does not expect structural case assignment to a subconstituent of a word.

In his study of the word in Eastern/Central Arrernte (a language spoken in the Alice Springs area, Australia), Henderson observes that in complex predicates of this language, "non-verbal morphemes can intervene between the two parts" (Henderson 2002: 114), as illustrated by the following example (in 5a, the complex predicate *arrernelheme* is split by the word *akewele* 'supposedly'):

(5) (a) arrerne akwele lh+eme
place SUPPO REFL+PRES
'supposedly sit down'
(b) arrern+elh+eme akwele
place+REFL+PRES SUPPO

Again, the complex predicates can be interpreted as forming lexical units that consist of two grammatical words, and hence they can be split without violating Lexical Integrity. Henderson (2002: 119) remarks that such combinations of a root and suffixes must sometimes be assumed to form one grammatical word because the root and the suffix part are co-dependent, as in the following example:

(6) apan+erle=arteke re ap+em+ele
feel+GO.ALONG$_1$=SEMBL 3SG:ERG GO.ALONG$_2$+PRES+SAME SUBJ
'like going along continuously feeling (its way)'

Henderson observes that "while the *–ap* part can be analyzed as a verb root, it does not occur productively as the sole root of a non-compound verb" (Henderson 2002: 119). This cannot to be taken as conclusive evidence for such complex words being one grammatical word. In Dutch, for instance, we find many particle verbs of which the verb does not occur on its own (e.g. *opkalefateren* 'to recover', of which the verb *kalefateren* does not occur on its own). Yet, we know for certain that such particle verbs form two grammatical words on the basis of both syntactic and morphological evidence (cf. Booij 2002). Note that there also syntactic idioms containing words that do not occur by themselves, but only in these idioms.

The phenomenon of word splitting is traditionally referred to as tmesis, and occurs in Ancient Greek and Classical Latin texts. However, the Greek and Latin verbs that are said to undergo tmesis can be analysed as cases of particle verb combinations as well, like those in Dutch discussed above. Hence, what is referred to as tmesis does not necessarily form counterevidence to the hypothesis that words cannot be split by syntactic rules.

Another phenomenon that is sometimes considered a case of tmesis is the insertion of a word between the syllables of another word, as in *abso-fuckin-lutely*. However, this phenomenon is a special kind of word formation that makes use of

the prosodic structure of its base words (prosodic morphology), and hence does not count against the claims that words cannot be split in the syntax.

2.1 Apparent Counterexamples to the Prohibition on Manipulation

Manipulation as referred to in the definition of Lexical Integrity in (2) above may have various forms: movement of parts of words to another position in the sentence (should be excluded, see the discussion of the examples (3 and 4) above), or the assignment or checking of morpho-syntactic properties of part of a word from outside, i.e. by a syntactic rule or constraint. Number agreement is an example of this latter kind of syntactic manipulation. The Principle of Lexical Integrity predicts that we will not find cases where the number of a word constituent is checked by number agreement. A well known form of number agreement in many languages is that between quantifier and head noun. If a quantifier expresses plural number, its head noun may have to appear with plural number as well. This is, for instance, the case for Spanish. However, it appears that both parts of Spanish copulative compounds have to be plural, as is the case for the copulative compound *poeta-pintor* 'poet-painter': its plural form is *poetas-pintores* (for example, *dos poetas-pintores* 'two poet-painters'.[2] This does not imply, however, that a syntactic rule of agreement of Spanish manipulates parts of words. Instead, this regularity can be interpreted as a morphological phenomenon: if we want to pluralize a Spanish copulative compound, both constituents must be marked for plural. Since Number agreement with *dos* requires a plural word, both parts of the compound are marked as plural.

The Hungarian example *tévét nez* 'be engaged in television watching' is also relevant here. As we saw above, the noun *tévé* is marked here as an accusative form. Since this word is not a word part, but a separate word that is part of a phrasal lexical unit, there is no manipulation of word-internal structure at stake here.

Another apparent counterexample to the 'no manipulation' part of the Lexical Integrity Principle is the phenomenon of gapping of parts of words. Gapping of parts of word is a clear case of manipulation of internal word structure. However, gapping is not necessarily a form of syntactic manipulation, and there is convincing evidence that it is prosodic in nature. Some examples from English and European Portuguese respectively are (Vigario 2003: 251):

(7) a. mono- and polysyllabic
inter- and intranational uses
homo- and heterosexual relations

[2] I thank Franz Rainer for bringing this case to my attention.

b. pré- e pós-guerra 'pre- and post-war'
segura- mas lentamente [= seguramente mas lentamente]
'surely but slowly'

(cf. Booij 1985 for a prosodic analysis of gapping in Dutch and German compounds). These are cases of prosodic gapping rather than syntactic gapping: a prosodic word is deleted under identity with another phonological word in the same phonological phrase. That is, this process of gapping does not refer to the morphological structure of words, but to their prosodic constituency, and operates at the level of prosodic structure. This is confirmed by the observation that this type of gapping is not restricted to structures with coordination. In the following Dutch and German examples, there is no coordination:

(8) a. Hij verwisselde de dagblad- voor de weekbladjournalistiek
He exchanged the newspaper- for the weekly-journalism
'He changed from newspaper journalism to weekly journalism'
b. Sie ersetzten Ofen- durch Zentralheizung
They exchanged stove- by central heating
'They changed from stove heating to central heating'

Therefore, this kind of gapping does not violate the principle of Lexical Integrity since this principle pertains to words in the syntactic sense, not to prosodic words.

The occurrence of inflected word forms as parts of complex words also deserves some discussion in this context, since at first sight the occurrence of such word-internal inflectional markings suggest that syntax can manipulate the internal structure of words. First, the inflection might be due to the fact that the relevant part of the complex word is a phrase, as is the case for the Dutch compound *oudemannenhuis* 'old men's home' discussed above. The well-formedness of phrases including their inflectional markings is determined by the syntactic module, even when they form parts of words, and hence this kind of manipulation does not count as counterevidence. Furthermore, we find inflected forms such as plural forms and case-marked forms of nouns in the non-head position of compounds (Booij 1994, 1996). As pointed out in these articles, the kind of inflection that we find there is inherent inflection that is determined by the semantics, and not by syntactic rules such as agreement or structural case marking that require the inflectional marking of word parts by a linguistic unit outside of that word. Hence, they form no counterevidence for the prohibition on syntactic manipulation of word-internal structure.[3]

[3] According to Bauer (2001: 704), "Booij's generalization [only inherent inflection feeds word formation] appears to be true, but it is probably not an absolute universal", because Bauer observed a number of cases in which finite verb forms are embedded in compounds, and a few cases of accusative marking of nouns within nominal compounds. Some of the examples that Bauer discusses, are not clear cases of compounding. This applies to the Hungarian lexical unit *tévé-t néz* discussed above in which the first part is marked as an accusative by means of the suffix *-t* (Bauer 2001: 704). As I mentioned above, there is clear evidence that such N V combinations are not compounds but phrasal in nature since the two parts can be split by other words such as the negative word *nem* 'not' (Kiefer 1992: 76).

2.2 Evidence for Accessibility

An example that suggests that access of syntactic or semantic rules to word-internal structure cannot be completely ruled out comes from Georgian. In Georgian we find expressions such as the following (Harris 2006a):

(9) sam tit-moč'r-il-i (k'aci)
three.OBL finger-cut.off-PTCP-NOM man.NOM
'(a man) with three fingers cut off'

The first word *sam* has to appear in the oblique form, because it modifies the word *tit* 'finger' which is part of the second word. That is, for the purpose of both case assignment (to the independent word *sam* only) and semantic interpretation, *sam* and *tit* form a unit. As Harris argues, the word *sam* cannot be considered a part of the next word (even though its form is indeterminate and could also be interpreted as a stem form), because recursive modification is not allowed within Georgian compounds. Hence, it should be interpreted as the oblique form of an independent word. This case assignment thus requires access to the internal morphological structure of the second word in (9). The construction in (9) may be compared to that in (10) where the first word bears nominative case, and you get a different interpretation:

(10) sam-i tit-moč'r-il-i
three-NOM finger-cut.off-PTCP-NOM
'three (men, people, statues) with fingers cut off'

In (10), the word form *sami* agrees in case marking with the second word as a whole, and hence it is a modifier of the whole word. Note that the word *tit*, being part of a compound, does not receive case marking itself.

This example reminds us of Corbett (1987), who also noted that the internal structure of complex words must sometimes be accessible for elements outside of that word. His prototypical example is the following Upper Sorbian phrase (Corbett 1987: 300):

(11) mojeho (gen. sg. masc.) bratrowe (nom.pl.) dźěći (nom.pl)
my brother's children

The selection of a gen.sg.masc. form of the possessive pronoun requires the pronoun selection process (agreement) to have access to the nominal base *bratr* 'brother' of the adjective *bratrowe*. The genitive case of *mojeho* depends for its semantic interpretation on the fact that the possessive pronoun and the nominal base *bratr* form a semantic unit. As in the Georgian example discussed above, it is not the case that a part of a complex word is manipulated by the syntax (in the sense that case is assigned to a word-internal constituent). Instead, the selection of a specific word form requires access to the internal structure of the complex word that it modifies. This once more underlines that manipulation and accessibility are different notions,

and should be distinguished when we investigate the viability of the principle of Lexical Integrity.

As far as the semantics is concerned, the Georgian example is similar to the English phrase *transformational grammarian* in which the adjective *transformational* modifies the constituent *grammar* of the complex word *grammarian.* This is a well known example of a bracketing paradox: semantically the adjective *transformational* forms a constituent with *grammar*, but syntactically it forms a unit with *grammarian*. A similar bracketing paradox can be observed for the English phrase *a hard worker* and its Dutch counterpart *een harde werker*. The word *hard* receives a specific adverbial interpretation 'with great intensity' which is dependent on the presence of a verb that it can modify. This verb is available as part of the deverbal noun *werk-er*. The Dutch version is even more interesting than the English one, because the Dutch word *hard-e* is overtly inflected as an adjective witness the presence of an inflectional ending *-e*.

The additional property of the Georgian construction is that the semantic scope of the numeral modifier includes the semantic information expressed by the case marking. Note once more that there is no case marking of the relevant word part itself. Hence, the possibility of manipulation of word-internal structure by the syntax can still be excluded, but access to the word-internal structure is necessary in order to give a proper interpretation to the semantic role of the word part *tit* and the oblique case of the numeral *sam*. We might therefore consider this example as showing that rules of semantic interpretation may need access to word-internal structure, a topic to be dealt with in more detail in the next section.

Another case of syntax requiring access to word-internal structure is the phenomenon of construction-dependent morphology. This is the situation in which a syntactic construction requires or allows a particular morphological form of words in that construction.

Normally, the syntax may require particular morphological properties (case, number, person, finiteness, etc.) to be present in words, as is the case for phenomena such as agreement and government, and the choice between finite or non-finite forms of verbs.

There are also construction-specific requirements of this sort. For instance, the Dutch progressive construction has the shape [*aan het* V-INF]$_{PP}$, as in *Ik ben aan het fiets-en* 'I am cycling' (Booij 2008). In that case, the syntax specifies information about the verb as a whole. It does not matter whether the infinitival form is realized by the suffix *-en*, the default suffix, or by *-n* (as is the case for a small set of Dutch verbs such as *doe-n* 'to do'). Moreover, the infinitival form of verbs is not tied to this construction, and is used in other constructions as well.

The form of construction-dependent morphology that is of special relevance for the discussion on lexical integrity is the situation in which words with a specific affix occur in a specific syntactic construction only. An example of that situation is the case of the inflection of Dutch numerals discussed in Booij (2005a). The use of the inflected forms of most numerals is restricted to a number of specific constructions that are exemplified in (12):

(12) a. *collective adverbial I*
met ons / jullie / hun drie-en
with us / you / their three-*en*
'the three of us /you / them together'

b. *collective adverbial II*
met zijn drie-en
with his three-*en*
'the three of us / you / them'

c. *temporal expressions*
bij zess-en
at six-*en*
'at about six o'clock'

na zeven-en
after seven-*en*
'after seven o'clock'

voor en-en
before one-*en*
'before one o'clock'

d. *number of parts*
Het schip brak in drie-en
The ship broke into three-*en*
'The ship broke into three pieces'

e. *appositive collective*
wij / ons drie-en
we / us three-*en*
'the three of us (SUBJ. / OBJ.)'

In present-day Dutch, the ending *-en* is used as one of the two plural suffixes for nouns (the other suffix is *-s*). Historically, however, the ending *-en* in the constructions (12) is a case ending, for instance the inflectional ending required by a preposition, as in (12a–c). The temporal expression *voor en-en* 'before one o'clock' (12c) makes it quite clear that *-en* cannot have been a plural suffix originally, since it would be semantically odd to use a plural form for the notion 'one hour'. The 1sg. plural pronoun *ons* in (12a) was originally the oblique form of a personal pronoun ('us') as required by the governing preposition, but could be reinterpreted as the possessive pronoun *ons* ('our'). Hence, the numerals in *-en* could be reinterpreted as plural endings in the collective constructions, and this is what happened (Van Loey 1959: 154). This is proven by (12e) where we find the numeral ending

in- *en* preceded by the subject, non-oblique form of the 1st person plural pronoun *wij* 'we'. This reinterpretation of the ending *-en* as a plural form is also confirmed by the change of the original form *twee-n* (the inflected form of *twee* 'two') into the plural form *twee-en*. This reinterpretation also led to the collective adverbial construction exemplified in (12b), with the fixed 3rd person singular possessive pronoun *zijn* 'his'. When this construction is used, there is no agreement for person and number with the subject, as illustrated by the following sentence:

(13) Wij gaan met zijn drie-en naar het feest
We go with his three-*en* to the party
'The three of us will go to the party'

In this sentence a 1.pl. subject is combined with the 3.sg possessive pronoun *zijn* 'his'. So there are two different collective adverbial constructions that are identical except that the possessive pronoun can be either a variable (and thus subject to the normal agreement constraints for possessive pronouns), or a fixed possessive pronoun *zijn* 'his'.

The reader will have noticed that in the glosses for the examples in (12) I did not make use of the morpho-syntactic feature PLURAL, but I mentioned the concrete Dutch suffix *-en* instead. There are two reasons for this. First, in (12c) an interpretation of *-en* as a plural suffix does not make sense semantically. Secondly, the specific suffix *-en* that is required by this construction cannot be equated with the abstract morpho-syntactic feature PLURAL, because this feature is expressed by either *-s* or *-en* depending on the prosodic make up of the stem (*−s* after an unstressed syllable, otherwise *-en*, cf, Booij 2002). For instance, the plural for the number 7 is *zeven-s* /zevəns/, as is the case when we talk about grading (*Jan kreeg twee zeven-s* 'John got two grades 7'). Yet, in the use of the word *zeven* shown in (12), the form of *zeven* is *zevenen* /zevənən/. The same applies to the number *negen* '9' /neɣən/: its plural is normally *negen-s*, but in these constructions it should be *negen-en*.

These observations imply that we have to specify the presence of a specific suffix *-en* in the constructions exemplified in (12). For instance, the constructional idiom for phrases like that in (12b) is:

(14) $[[\text{met}]_{P}\ [\text{z'n}\ [[\text{x}]_{\text{Numeral}}\ \text{-en}]_{N}]_{NP}]_{PP}$
with his x-en
'the x of us / you / them'

This analysis implies that the principle of Lexical Integrity as formulated in (2) is too strong, and that the syntax may require access to the internal morphological structure of words. This admittedly special situation is the effect of the rise of syntactic constructions in which specific morphological information is 'frozen'.

In short, syntactic constructions may require the presence of specific morphemes to be present in words, and hence the visibility of word-internal structure to syntax cannot be excluded completely.

3 Accessibility of Word-Internal Structure: Semantic Scope Phenomena

The discussion in the preceding section implies that, although syntax may not be allowed to manipulate the internal structure of word, there are cases in which syntax does require access to the internal structure of words. This means that the interface between morphology and syntax is such that the syntax may have to see word-internal morphological structure.

Word-internal structure needs to be visible to phonology as well. There is abundant evidence that the computation of the correct phonetic form of complex words may require information about morphological structure, This holds in particular for the computation of prosodic forms of words (syllabification and stress patterns) (Booij 2005c).

A-morphous morphology, as defended in Anderson (1992) for the domain of derivation – not for compounding which he considers as being syntax-like - , has also to be rejected for morphological reasons. There is massive evidence that the internal structure of complex words has to be accessible to morphological processes (cf. Carstais-McCarthy 1994, Booij 2002).

This position is corroborated by psycholinguistic findings. Psycholinguistic research has provided clear evidence for the existence of word families, i.e. families of words that share one or more constituents (Schreuder and Baayen 1997). The notion 'word family' presupposes the accessibility of word-internal structure as well, because a family is to be defined as a set of words that share one or more morphological constituents. Hence, the size of a family can only be determined if the word-internal structure of words is accessible.

In this section I will discuss some data on the accessibility of word-internal structure of complex word to another level of the grammar, that of semantics. In particular, in Dutch NPs of the form $[(\text{Det})\ A + N]_{NP}$, the adjective may have scope over the first constituent of the complex word only, rather than over the complex word as a whole. The occurrence of this kind of restricted scope is pervasive in texts on issues of government policy. My source of data is Bijker and Peperkamp (2002), a Dutch science policy document with a title that translates as *Involved humanities. Perspectives on cultural changes in an era of digitalization.* The following phrases can be found in this document:

(15) $[A\ [NN]_N]_{NP}$

a. visuele informatie-verwerking
visual information processing
'processing of visual information'

b. intellectuele eigendoms-rechten
intellectual property rights
'rights of intellectual property'

c. taalpolitieke beleids-makers
language political policy makers
'language policy makers'

d. electronische reproductie-rechten
electronic reproduction rights
'rights of electronic reproduction'

e. digitale kennis-omgevingen
digital knowledge environments
'environments of digital knowledge'

f. wetenschappelijke kennis-cyclus
scientific knowledge cycle
'cycle of scientific knowledge'

g. publieke oordeels-vorming
public opinion formation
'formation of public opinion'

h. interactieve gebruiksmogelijkheden
interactive use possibilities
'possibilities for interactive use'

(16) [A [N-suffix]$_N$]$_{NP}$

wetenschappelijke onderzoek-er
scientific research-er
'person who does scientific research'

(17) [A [A-suffix]$_N$]$_{NP}$

a. wetenschappelijke deskundig-heid
scientific expert-ness
'scientific expertise'

b. digitale vaardig-heid
digital competent-ness
'digital competence'

In these examples, the first word, an inflected adjective, has semantic scope over the first part of the second (complex) word. For instance, *visuele informatieverwerking* 'visual information processing' is the processing of visual information, not the visual processing of information. Similarly, in the phrase *wetenschappelijke deskundigheid*, the adjective *wetenschappelijke* 'scientific' modifies the part

deskundig 'competent': it refers to the property of being a scientific expert, not to the scientific property of being an expert. All these examples are therefore 'semantic bracketing paradoxes' of the same type as *transformational grammarian* (cf. Spencer 1988 and Beard 1991 for discussion).

It is obvious that these are not cases where the first two words form a phrasal constituent that is embedded in a compound, given the inflectional rules for Dutch adjectives. For instance, if the phrase *intellectuele eigendomsrechten* had the structure $[[intellectuele\ eigendoms]_{NP} rechten]_N$, the presence of the inflectional ending *–e* at the end of *intellectuele* could not be explained, since the correct phrase is *intellectueel eigendom*, without the final inflectional *–e*, because *eigendom* is a neuter noun that requires *intellectueel* as the attributive form of the adjective.

In Spencer's (1988) analysis, the restricted scope interpretation of *transformational grammarian* is related to the existence of a lexical unit *transformational grammar* and the word pair *grammar-grammarian*, and therefore analysed as an analogical formation. However, restricted scope also occurs in cases where such an explanation cannot be invoked. In the examples if restricted semantic scope given above such NPs are not available. For instance, there is no well formed phrase *digitale vaardig* that can be invoked as a basis for the forming of *digitale vaardig-heid*, since the relevant AP should have the form *digitaal vaardig* 'digitally competent', without the first word being inflected: it is an adverb, not an adjective in pre-adjectival position. Similarly, there is not a well-formed phrase *wetenschappelijke onderzoek* 'scientific research' that can be related to *wetenschappelijke onderzoeker* 'scientific researcher' since the correct phrase is *wetenschappelijk onderzoek,* without a final schwa on the adjective.

These facts clearly show that word-internal structure must be visible to rules of semantic interpretation. Hence, the Principle of Lexical Integrity should be phrased in such a way that it does not forbid rules for the semantic interpretation of phrasal constituents to have access to word-internal structure.

The accessibility of word-internal structure for reasons of semantic interpetation is also important at the word level, for the proper scope assignment of bound modifiers such as prefixes within complex words. This is illustrated by the following two Dutch examples:

(18) pro-Pakistaan-s-e extremisten 'pro-Pakistan extremists'
tussen-gemeente-lijk-e oplossingen 'inter-council solutions'

In the first example the part *Pakistaans* is an adjective derived from the noun *Pakistan* (the *-e* is an inflectional ending). The semantic scope of *pro* is the nominal base *Pakistan* as shown by the interpretation specified in the gloss. In the second example we see a denominal adjective *gemeente-lijk* derived from the noun *gemeente* 'council'. Clearly, the prefix *tussen* 'between' must have scope over the nominal part *gemeente* only, given the meaning 'between councils' of the adjective *tussen-gemeentelijk*. This observation, however, is not so much an argument concerning Lexical Integrity as rather one against a-morphous morphology.

4 Lexical Integrity and Construction Morphology

The restriction of Lexical Integrity to a prohibition on movement, splitting, deletion of parts of words and the assignment by the syntax of morpho-syntactic properties to word parts, receives a natural explanation in the framework of Construction Grammar and the constructionist view of morphology defended in Booij (2005b) for word formation; (cf. Blevins 2006, and Gurevich 2006 for an application of the insights of Construction Grammar to the domain of inflection).

> "In Construction Grammar, the grammar represents an inventory of form-meaning-function complexes, in which words are distinguished from grammatical constructions only with regard to their internal complexity. The inventory of constructions is not unstructured; it is more like a map than a shopping list. Elements in this inventory are related through inheritance hierarchies, containing more or less general patterns." (Michaelis and Lambrecht 1996: 216)

Word can be seen as constructions on the word level. Grammaticalization is an essential factor in the historical development of morphology from syntactic structures, and hence it should come as no surprise that complex words have constructional properties: an internal structure that is visible to rules of the grammar. One of the best known examples in this respect is the rise of compounds from phrasal patterns. Therefore. it has often been stated in the literature that compounds are still syntax-like, and not always easy to demarcate from phrases (cf. Dahl 2004: Chapter 10 for extensive illustration and discussion of the continuum syntactic construction – compound word).

The Dutch examples of construction-dependent morphology discussed in Booij (2005a) are all cases where a specific construction preserved morphological structure that is no longer regular from a synchronic point of view and has therefore to be specified as part of the construction. Such forms of accessibility of morphological information to syntax are to be expected if the notion 'construction' has a central role in the grammar of natural languages.

The main reason why we consider a sequence of morphemes a word is that that sequence behaves as a cohesive unit with respect to syntactic processes. In other words, cohesiveness is the defining criterion for canonical wordhood, whereas other properties such as being a listeme (a conventional expression) are clearly not to be seen as defining properties for wordhood. Hence, if we take the notion word seriously, we might say that its defining property is cohesiveness or non-interruptability. In other words, the 'no manipulation' part of the principle of Lexical Integrity is the proper interpretation of word cohesiveness.

The constructionist view of morphology does not imply that morphology can be equated with the syntax of morphemes. The word remains an essential unit for stating regularities. Both words and phrasal constructions are domains over which certain generalizations can be stated, and hence the domains of 'word' and 'phrase' are both essential for the analysis of natural languages (Blevins 2006).

5 Conclusions and Discussion

The conclusion of this paper is that the principle of Lexical Integrity has to be formulated in such a way that it does not exclude the accessibility of word-internal structure. The question then remains whether we can formulate this principle as an absolute universal that forbids syntactic manipulation.

In this restricted form, this principle is a further interpretation of the universal principle that all languages distinguish between words and phrases. The distinction between complex words and phrases makes only sense if the word exhibits a higher degree of cohesiveness than the phrase, and hence we need the 'no syntactic manipulation' constraint to give substance to the distinction between words and phrases.

Although this form of lexical integrity seems to be the default situation for natural languages and thus defines the canonical notion 'word', there are exceptional cases in which even this restricted form of lexical integrity appears to be a violable constraint.

An example is the behaviour of endoclitics in Udi, a North-East Caucasian language. The relevant facts are discussed in Harris (2000): clitics that function as person markers can appear word-internally, that is as endoclitics, in between two morphemes in complex verbs (and morpheme-internally in simplex verbs).[4,5]

Another relevant case is that of Arrernte, a language discussed above in Section 2. Henderson reports that in this language we may have 'initial separation': "the first two, or rarely three syllables of a verb can optionally be separated from the remainder of the verb. Intervening material seems to be limited to particles, clitics, pronouns, and simple NPs" Henderson (2002: 121). The most telling example is that in which a lexical root 'to cough' is split into two parts due to the presence of an intervening word:

(19) ateke akwele tn+eme
cough$_1$ SUPPO cough$_2$+PRES
'(she's) supposedly coughing'

Hence, there appear to be cases of real violation of lexical integrity, although this is cross-linguistically a marginal phenomenon.

Finally, it should be mentioned that there are also languages in which the internal structure of complex words is accessible to rules of anaphora (Harris 2006b).

In conclusion, this paper has shown that the principle of Lexical Integrity should be formulated in such a way as not to exclude the different modules of the grammar from ever having access to word-internal structure. Moreover, Lexical Integrity as the prohibition on syntactic manipulation of word-internal constituents is not an absolute universal, but rather the default situation.

---

[4] A similar case of clitic intrusion is reported for Sorani Kurdish in Samvelian (2006).

[5] Therefore, Anderson (2005: 161–165) concluded that Lexical Integrity is a violable constraint in the sense of Optimality Theory. In Anderson's analysis, this constraint is normally undominated, but in Udi, the positional constraints on person marker clitics are ranked higher, and hence, such clitics can appear word-internally, thus behaving as endoclitics.

References

Ackema, Peter and Ad Neeleman. 2005. *Beyond Morphology. Interface Conditions on Word Formation*. Oxford: Oxford University Press.

Anderson, Stephen R. 1992. *A-morphous Morphology*. Cambridge: Cambridge University Press.

Anderson, Stephen R. 2004. Morphological universals and diachrony. In *Yearbook of Morphology 2004*, eds. Geert Booij and Jaap van Marle 1–18. Dordrecht: Springer.

Anderson, Stephen R. 2005. *Aspects of a Theory of Clitics*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2001. Compounding. In *Language Typology and Language Universals*, eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, Vol. 1, 695–707. Berlin: De Gruyter.

Beard, Robert. 1991. Decompositional composition: The semantics of scope ambiguities and "bracketing paradoxes". *Natural Language and Linguistic Theory* 9: 195–229.

Bijker, Wiebe and Ben Peperkamp. 2002. *Geëngageerde geesteswetenschappen. Perspectieven op cultuurveranderingen in een digitaliserend tijdperk*. Den Haag: Adviesraad voor Wetenschaps- en Technologiebeleid.

Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42: 531–573.

Blom, Corrien. 2005. *Complex predicates in Dutch, Synchrony and Diachrony*. Ph. D. Diss. Vrije Universiteit Amsterdam. Utrecht: LOT.

Booij, Geert. 1985. Conjunction reduction in complex words: a case for prosodic phonology. In *Advances in Non-linear Phonology*, eds. Harry van der Hulst and Norval Smith, 143–160. Dordrecht: Foris.

Booij, Geert. 1994. Against split morphology. In *Yearbook of Morphology 1993*, eds. Geert Booij and Jaap van Marle, 27–50. Dordrecht: Kluwer.

Booij, Geert. 1996. Inherent versus contextual inflection and the Split Morphology hypothesis. In *Yearbook of Morphology 1995*, eds. Geert Booij and Jaap van Marle, 1–16. Dordrecht: Kluwer.

Booij, Geert. 2002. *The Morphology of Dutch*. Oxford: Oxford University Press.

Booij, Geert. 2005a. Construction-dependent morphology. *Lingue e Linguaggio* 5: 163–178.

Booij, Geert. 2005b. The demarcation of derivation and compounding: evidence for Construction Morphology. In *Morphology and its Boundaries*, eds. Wolfgang U. Dressler, Dieter Kastovsky, Oskar Pfeiffer, and Franz Rainer, 111–132. Amsterdam / Philadelphia: John Benjamins.

Booij, Geert. 2005c. The interface between morphology and phonology. *Skase Journal of Theoretical Linguistics* 2: 17–25.

Booij, Geert. 2008. Constructional idioms as products of linguistic change: the *aan het* + INFINTIVE construction in Dutch. In *Construction Grammar and Language Change*, eds. Alexander Bergs, and Gabriele Diewald, 79–104. Berlin: Mouton de Gruyter.

Bresnan, Joan and Sam A. Mchombo. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13: 181–254.

Bybee, Joan. 1985. *Morphology. A Study of the Relation between Meaning and Form.* Amsterdam / Philadelphia: John Benjamins.

Bybee, Joan. 2006. Language change and universals. In *Linguistic Universals*, eds. Ricardo Mairal, and Juana Gil, 179–194. Cambridge: Cambridge University Press.

Bybee, Joan L., Revere Perkins, and William Pagliuca. 1994. *The Evolution of Grammar. Tense, Aspect, and Modality in the Languages of the World*. Chicago: Chicago University Press.

Carstairs-McCarthy, Andrew. 1994. Morphology without word-internal constituents A review of Anderson (1992). In *Yearbook of Morphology 1993,* eds. Geert Booij and Jaap van Marle, 209–234. Dordrecht: Kluwer.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Comrie, Bernard. 2001. Different views of language typology. In *Language Typology and Language Universals,* eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible, Vol. 1, 25–39 Berlin: De Gruyter.

Corbett, Greville G. 1987. The morphology/syntax interface. *Language* 63: 299–345.
Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam / Philadelphia: John Benjamins.
Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language*, ed. Joseph Greenberg, 73–113. Cambridge Massachusetts MIT Press [1966^2].
Gurevich, Olga I. 2006. *Constructional Morphology. The Georgian Version.* Ph. D. diss, Univ. of California at Berkeley.
Hawkins, John A., and Anne Cutler. 1988. Psycholinguistic factors in morphological asymmetry. In *Explaining Language Universals,* ed. John Hawkins, 280–305. Oxford: Blackwell.
Harris, Alice. 2000. Where in the word is the Udi clitic? *Language* 76: 593–616.
Harris, Alice. 2004. The challenge of typologically unusual structures. In *Morphology and Linguistic Typology. On-line Proceedings of the Fourth Mediterranean Morphology Meeting, Catania, 21–23 September 2004,* eds. Geert Booij, Emiliano Guevara, Angela Ralli, and Salvatore Sgroi. http://mmm.lingue.unibo.it/proc-mmm4.php
Harris, Alice. 2006a. In other words: External modifiers in Georgian. *Morphology* 16: 205–229.
Harris, Alice. 2006b. Revisiting anaphoric islands. *Language* 82: 114–130.
Haspelmath, Martin. 2006. Creating economical morphosyntactic patterns in language change. To appear in *Language Universals and Language Change*, ed. Jeff Good. Oxford: Oxford University Press.
Haspelmath, Martin. 2007. Explaining some universals of causative verb formation. Lecture, LSA meeting Anaheim.
Helmbrecht, Johannes. 2004. Cross-linguistic generalizations and their explanations. In *Morphology. An International Handbook on Inflection and Word Formation*. Vol. 2, eds. Geert Booij, Christian Lehmann, Joachim Mugdan, and Stavros Skopeteas, 1247–1254. Berlin: De Gruyter.
Henderson, John. 2002. The word in Eastern / Central Arrernte. In *Word. A Cross-linguistic Typology*, eds. R. M. W. Dixon and Alexandra Y. Aikhenvald, 100–124. Cambridge: Cambridge University Press.
Kiefer, Ferenc. 1992. Compounding in Hungarian. *Rivista di Linguistica* 4: 61–78.
Lieber, Rochelle, and Sergio Scalise. 2006. The Lexical Integrity Hypothesis in a new theoretical universe. *Lingue e Linguaggio* 6: 7–32.
Loey, A. van 1959. *Schönfelds Historische Grammatica van het Nederlands.* Zutphen: Thieme.
Michaelis, Laura A., and Knud Lambrecht 1996. Toward a construction-based theory of language function: The case of nominal extraposition. *Language* 72: 215–247.
Samvelian, P. 2006. Pronominal clitics in Sorani (Central) Kurdish Dialects. Paper presented at the 5th Décembrettes Morphology Meeting, Université de Toulouse, 7–8 December 2006.
Scalise, Sergio (ed.). 1992. *The Morphology of Compounding*. *Rivista di Linguistica* 4: 1–251.
Schreuder, Robert and Harald Baayen. 1997. How complex simplex words can be. *Journal of Memory and Language* 37: 118–139.
Spencer, Andrew. 1988. "Bracketing paradoxes" and the English lexicon. *Language* 64: 663–682.
Vigario, Marina. 2003. *The Prosodic Word in European Portuguese*. Berlin: Mouton de Gruyter.

Searching for Universals in Compounding

Emiliano Guevara and Sergio Scalise[1]

Abstract Despite the fact that compounding is the most widespread word-formation strategy in the world's languages – and for some languages, the only one – compounds have received little, if any attention in linguistic typology and, in particular, in studies on linguistic universals.

This paper presents a preliminary cross-linguistic overview of the basic features of compounding, their distribution and expected degree of variation on the basis of the Morbo/Comp database. A basic conceptual framework of four important points for the study of compounding from a typological point of view is described, including the definition and classification of compounds, the position of the head constituent and the definition of compound type. These issues necessarily have to be dealt with prior to the actual examination of the data, for they play a fundamental role in the design and implementation of the research methodology.

We attempt to extrapolate universal tendencies in compounding with respect to the following list of characteristic features: output categories, input categories, the relation between the constituents, headedness and the combination of categories. A series of universal scales of preference describing the world's languages behaviour in the domain of compound formation is proposed. Some well-known putative universals in compounding are discussed and rejected on the basis of our data, namely whether compounds are only right-headed, the root-compounding parameter and the head as *locus inflectionis*.

Keywords Universals · compounding · word-formation · classification of compounds · typological database

E. Guevara
University of Bologna

S. Scalise
University of Bologna

[1] The research underlying this paper has been funded by the Italian Ministry of University and Research (project CompoNet, PRIN 2005). While the authors have discussed together the whole paper, S. Scalise is responsible for Sections 1, 2, 3, 4.2 and 4.3, and E. Guevara is responsible for Sections 4.1, 4.4, 5, 6 and 7. We thank G. Booij, B. Basciano, A. Ceccagno, A. Fábregas and E. Vineis for their comments on previous versions of this paper. We also thank two anonymous reviewers for their helpful and constructive observations.

S. Scalise et al. (eds.), *Universals of Language Today*, 101–128,
DOI: 10.1007/978-1-4020-8825-4_6,

There are probably no languages without either compounding, affixing, or both. In other words, there are probably no purely isolating languages. There are a considerable number of languages without inflections, perhaps none without compounding and derivation.

(J. Greenberg 1963: 92)

1 Introduction

The target of this work is to propose lines of empirical research and methodological suggestions towards the study of universals in compounding on the basis of the Morbo/Comp project.

Compounding is a rather neglected phenomenon in typological studies:[2] there is no trace of compounding even in the best repertories of Universals (e.g. *The Universals Archive*, F. Plank, Konstanz) and very little attention to compounds is given in typological works (e.g. *World Atlas of Linguistic Structures*, Haspelmath, Dryer, Gil, Comrie 2005). This is somewhat surprising since – as it is well known – compounds are the morphological constructions which are closest to syntactic constructions, to the point that it is sometimes difficult to distinguish between compounds and phrases. Therefore, a complete database of syntactic Universals should, in our view, also include compounding.

In this paper we will first introduce, very briefly, the Morbo/Comp project (cf. Section 2 below), which is the source of our data. We will then present a short speculation on the universal nature of compounding (cf. Section 3). Next, we will discuss four delicate and fundamental issues in compounding (cf. Section 4), namely, the definition of compound (cf. Section 4.1), the classification of compounds (cf. Section 4.2), the notion of "head" and the position of the head-constituent (cf. Section 4.3) and the identification of compound-types (cf. Section 4.4). We will then apply this framework to the currently available compounding data in the Morbo/Comp database (cf. Section 5). We will attempt to extrapolate universal tendencies in compounding as far as the following list of characteristic features are concerned: (a) output categories (cf. Section 5.2), (b) input categories (cf. Section 5.3), (c) the relation between the constituents (cf. Section 5.4), (d) headedness (cf. Section 5.5), (e) combination of categories (cf. Section 5.6). Next, we will discuss some false universals in compounding, namely: (a) whether compounds are only right-headed (cf. Section 6.1), (b) the root-compounding parameter

[2] There is a long standing tradition of studies devoted to compounding in theoretical linguistics, which is traceable back to Bloomfield (1933) and Lees (1960) with respect to the purely formal aspects of compound formation, and to Levi (1978) for the aspects of semantic compositionality in compounding. However, there is no explicit mention of possible Universals in compounding. See, for example, Spencer (2006: 129), where a rather pessimistic view emerges: "If we think of morphology as the study of word structure, we are greatly hampered by the fact that we have no really good understanding of what could constitute a universal characterization of morphological wordhood".

(cf. Section 6.2), (c) the head as *locus inflectionis* (cf. Section 6.3). Finally, we will draw some preliminary conclusions (cf. Section 7).

2 The Morbo/Comp Project

The final target of Morbo/Comp[3] is a database of compounds, covering the six macro-areas proposed by Dryer (1992) (Africa, Eurasia, South East Asia and Oceania, Australia-New Guinea, North America, South America). The major source of inspiration for Morbo/Comp comes from the experience gathered in previous large data-driven typological projects such as Matthew Dryer's Typological Database, the Surrey Morphology Group's databases and Bauer's (2001) work on compounding. Among these, Bauer's work is the only one specifically dedicated to compounding.

Bauer (2001) analyzes 36 different languages, six for each of Dryer's macro-areas. Data are taken from existing works, dictionaries and grammars. Since in our project compounds are analyzed in 18 different fields (as it will be clarified below) requiring thus a subtle analysis for each compound, we could not possibly rely on grammars and dictionaries only. For such an analysis, one needs all data checked by a native speaker linguist. This drove us to analyse compounds of those languages for which we could use the judgement of a native speaker (with some obvious exceptions such as Latin), and therefore our sample is not yet well balanced.

At present, we have, thus, ∼80.000 compounds for 21 languages.[4] In this database, the Romance, Germanic and Slavic families are currently well represented. Then there are also three languages from an "East Asian" group, two Finno-Ugric languages, Latin and Modern Greek. Data sources for the project are specific studies, lexical databases (when available), competence of native speakers, dictionaries and grammars.[5] The sample we use in this paper is a selected subset comprising ∼3000 compounds in 16 different languages (Catalan, French, Italian, Spanish, Dutch, English, German, Norwegian, Swedish, Bulgarian, Polish, Russian, Serbo-Croatian, Chinese, Japanese and Korean), each compound being analyzed in 18 searchable fields.[6]

---

[3] Morbo/Comp stands for 'Morphology at Bologna Research Group/Compounding'.

[4] Bulgarian (I. Krapova), Catalan (E. Bernal), Chinese (A. Ceccagno & B. Basciano), Dutch (E. Guevara), English (A. Bisetto), Finnish (J. Niemi), French (F. Villoing), German (A. Luedeling & P. Adolphs), Modern Greek (A. Ralli), Hungarian (T. Toldi), Italian (S. Scalise), Japanese (F. Forza), Korean (data from Sohn 1999), Latin (R. Oniga), Norwegian (L. Del Ponte), Polish (B. Szymanek), Russian (C. Melloni), Serbo-Croatian (S. Rakic), Spanish (E. Guevara), Swedish (S. Niemi), Turkish (A. Goksel & B. Haznedar). Notice, however, that Morbo/Comp is an open project and new languages can be added at any moment.

[5] Furthermore, semi-automatic extraction and classification from corpora is currently under development for Italian (Baroni, Guevara, Pirrelli, 2007).

[6] Language, Compound, Output category, Internal structure, Classification, Categorial head, Semantic head, 1st constituent, Category of the first constituent, Linking element, 2nd constituent, Category of the second constituent, Linking element 2, third constituent, Category of third constituent, Plural, Gender, Gloss and translation, Observations.

3 Universality of Compounds?

The long-standing debate between formal and functional linguists regarding the search of universals in human language is far from being settled. However, the two viewpoints are not totally incompatible since it is today maintained that they can be viewed as complementary approaches (cf. the recent reviews in Mairal and Gil 2006 and Croft, this volume).

In this paper we will not defend either position also because there are still very pessimistic opinions on the very possibility that true language universals exist (universals intended as structures that can be found in every language, cf. for example Croft 2004, Tomasello 2003, 2006).

However, sceptical views regarding the existence of true universals do not imply that universals do not exist altogether. As claimed by Tomasello (2003: 5) universals are not to be found in particular linguistic constructions, but rather must be looked for in different, higher-order levels of analysis. Two possible places to host universals are general macro-concepts such as *human cognition* (Croft 2001, Talmy 2003) or *communication* (Comrie 2003).

This point of view can be fruitfully applied to compounding: one could maintain, for example, that compounds fulfil a communicative strategy that is intrinsically different from that of syntactic expressions and, therefore, their existence could be motivated by human communication purposes.

In particular, compounding realises the tendency towards multiword constructions such as idioms, collocations, binomial constructions or the so-called prefabs. Furthermore, compounds also serve the function of effectively compressing the information that is contained in an utterance. Besides allowing data-compression, compounds also offer a rich source of metaphoricity: metaphoric expressions, far from being a limited and circumscribed phenomenon, are pervasive in human language (witness common examples such as *snail-mail* or *cat-fish*, cf., among many others, Gottfried 1997). Another possible reason for the productivity of compounding is the flexible nature of the semantic relation between the constituents, especially so in NN compounds.[7]

Furthermore, the grammatical relations between the constituents of a compound (as it will be clarified in Section 4.2) are not exclusive to compounding: they are shared, *mutatis mutandis*, also by syntactic constructions, and as such, they are presumably universal.

Other aspects of compounding are instead better handled by formal theoretical approaches. The standard position adopted in Chomskyan linguistics is that Universals cannot be found in any construction whatsoever, but that rather they must be looked for in the general operations that make it possible for any language to build any kind of construction.[8] All languages equally share the capacity of e.g. merging

---

[7] Cf. Levi (1978).

[8] Cf. the recent overview of the framework in Chomsky (2005) and the interesting discussion in Hauser, Chomsky & Fitch (2002), Pinker & Jackendoff (2005), Fitch, Hauser & Chomsky (2005) and Jackendoff & Pinker (2005).

two constituents together, moving constituents from and to certain positions in a structural tree, determining which constituent is going to project in a structure, etc.

For example, the position of the head-constituent in compounding could be incorporated straightforwardly in the Principles and Parameters framework, as a candidate feature for Universal Grammar.[9] And this standpoint could be held true even if compounding were proven not to be attested in every language. As Jackendoff (2002) puts it:

> "[…] non-universal aspects of linguistic structure may be candidates for Universal Grammar as well. When you have a toolkit, you are not obliged to use every tool for every job. Thus we might expect that not every grammatical mechanism provided by Universal Grammar appears in every language."
>
> (Jackendoff 2002: 75)

The preceding discussion leads us to think that it is not pointless to look for universal properties in compounding, even though, as maintained by Haspelmath (2007), there are no "pre-established categories for language description" and, as a consequence, the job a linguist must do "is to describe the phenomena in as much detail as possible". Our goal in this paper is, thus, pre-theoretic: to accurately describe the basic features of compounding, their distribution and expected degree of variation.

4 Important Issues for the Typological Study of Compounding

In the next subsections we will discuss four important points for the study of compounding, especially from a typological/universal point of view: the definition of compound, the classification of compounds, the position of the head constituent and the definition of compound type. These issues necessarily have to be dealt with prior to the actual examination of the data, for they play a fundamental role in the design and implementation of the research methodology: together, they constitute what could be called a "prolegomenon" to the study of compounds. Despite the still tentative nature of our exposition, we believe that the points at stake should be taken into consideration in any serious treatment of compounding. It will be immediately clear that our matter of concern cannot be tackled empirically without a previous conceptual "pre-processing". In order to compare disparate data from the world's languages, a basic framework must be defined: at present, we can only provide a tentative sketch of the whole framework that, however, enables us to begin a typological study of compounding on the basis of a large cross-linguistic set of data.

[9] However, recent work on Chinese compounds seems to contradict the idea that the position of the head-constituent in compounding is a binary parameter (cf. Ceccagno & Basciano, 2007).

4.1 Definition of "Compound"

A common procedure in scientific practice is trying to spell out a clear definition of the object of study. Unfortunately, the definition of 'compound' is a particularly difficult task.

The existing definitions are neither totally satisfactory nor sufficiently comprehensive. Consider for example the following:

(1) "[...] A lexical unit made up of two or more elements, each of which can function as a lexeme independent of the other(s) in other context, and which shows some phonological and/or grammatical isolation from normal syntactic usage" (Bauer 2001)

Very clearly, this definition cannot be adequate, because in turn various subparts of it are too vague. For instance, we do not know how to precisely identify notions such as "*some phonological and/or grammatical isolation*" just as we do not understand how "*normal syntactic usage*" can be defined. Consider now the following definition:

(2) "Composition [...] denotes the combining of two free forms or stems to form a new complex word referred to as compound. [...] Composition has come to be viewed in current linguistic work as the process of concatenating two existing stems from the lexicon of a language to form a new, more complex stem which has the potential to enter the lexicon as a stable morphological unit." (Olsen 2000)

This tentative definition, although more precise than the previous one, still suffers from a series of problems: for example, it takes for granted that the constituents in a compound are *stems*. While this is true for some languages, it does not hold for other languages that allow roots or fully inflected words to take part in a compound. In some cases, even whole phrases or sentences are said to be part of a compound (cf. *She had a [devil-may-care approach]*, *This is a [God-is-dead theology]*). Another problem is the fact that not all compounds productively created actually enter the lexicon: often a compound serves the function of a *nonce word*, ending its "life" right after it has been created; this fact, however, does not ban such compounds from taking part in larger structures, such as (multiply) recursive *nonce* compounds, without necessarily having to be lexically listed as a requirement.

Generalizing, the many definitions of *compound* that one finds in the literature are tightly predetermined by the theoretical choices made by the author(s). Consequently, one's views and beliefs regarding the fundamental notions of morphology – and of linguistics in general – are critical in shaping a working definition for compounding. In other terms, one's conception of hotly debated (and never agreed upon) issues such as *word*, *morpheme*, *stem*, *root*, *lexicon*, *concatenation*, etc., will contribute in shaping one's definition of *compound*. The consequences can even reach an absurd limit: if a theoretical approach considers compounds to be mere syntactic constructs, there will be simply no possible definition to propose (other than saying that compounds are some sort of phrasal construction).

It remains to prove whether any existing theoretical approach is useful and applicable in order to draw a typological/universal sketch of compounding in the world's languages.[10]

Thus, we will not try to propose yet another definition of *compound*. There are in fact many fundamental notions in linguistics that are ill-defined but, nevertheless, constantly used in the literature in an intuitive way (cf. *sentence*, *phrase*, *word*, etc.), and compounding may very well be one of them.[11] We can simply set aside the problem of the definition of compound and start looking for general tendencies in the world's languages.

A good starting point is to ask ourselves what is special about compounds, what makes a compound different from, e.g., a syntactic construction. The first possible answer is that compounds are special (with respect to the corresponding phrases) because in a compound there are typically two constituents that are held together by a relation which is not explicitly (phonetically) realized: for example, *apron string* can be paraphrased (and is probably interpreted) as *the string **of an** apron*.

The inner essence of a compound can be captured (in the prototypical case) with the following rough schema, where X, Y and Z represent major lexical categories, and $\Re$ represents an implicit relationship between the constituents (a relationship not spelled out by any lexical item):

(3) [X $\Re$ Y]Z

The problem that arises from (3) is how to establish the nature of $\Re$. For example, Bisetto and Scalise (2005) claim that $\Re$ is in the first place a *grammatical relation* (an idea which can be traced back to Bloomfield 1933 and Marchand 1969). If we consider two compounds such as *apron string* and *poet painter* we can readily notice that the grammatical relation that holds between the constituents in each of them is not the same: their corresponding paraphrases are very different, 'string of **an/the** apron' vs. 'poet **and** painter'. However, there is still no certain answer with respect to the nature of $\Re$: besides the traditional view holding that $\Re$ is a *grammatical* relation, other approaches suggest a semantic nature for $\Re$ based on the processes of concept-combination (cf. for instance Wisniewski and Gentner 1991, Wisniewski 1997, Gagné and Shoben 1997, Costello and Keane 2000, etc.).

A second starting point for our research is the consideration of what a *canonical* or *prototypical compound* turns out to be in the languages of the world. This idea is very much in agreement with the so-called canonical approach to linguistic typology, advocated by G. Corbett and the Surrey Morphology Group (cf. Corbett, this volume; Corbett, 2005). The canonical approach seems to be particularly appropriate for the construction of database-resources and their use in typological research.

---

[10] Notwithstanding the fact that the choice of a theoretical framework is crucial for the full understanding of compounding, we will not defend here any position. However, as for the main target of the present paper, we believe that our data and analysis are fully compatible with recent theoretical proposals such as Ackema & Neeleman (2004), Jackendoff (1997, To appear) or Di Sciullo (2005).

[11] Notice that the notion *compound* stands halfway between *word* and *phrase*.

The *canonical* instance in compounding can be thought of as a multiword expression that realises the intersection of (at least) the following set of converging criteria:[12]

(4) a. syntactic atomicity (no anaphoric relations between an internal constituent of a compound and an external element[13]);
b. lexical integrity[14];
c. lexical nature of constituents (lexemes, i.e. words, stems or roots), members of one of the major lexical categories;
d. the whole is a member of one of the major lexical categories.

Thus, the *canonical* in compounding seems to match quite closely the most productive compound-types of well-studied languages, such as Germanic, Romance or Chinese. This does not come by chance: a great part of the literature on compounding has somewhat tacitly adopted the canonical approach. The *canonical* instance in compounding is exemplified by the best-known patterns, those that have attracted the most attention.

However, by the canonical approach it is expected that the languages of the world may present compound-types that stand apart from the simplest and most familiar instances. To take but one example, consider the so-called phrasal compounds, which diverge from the *canonical* in that one of the constituents is syntactic, not lexical, in nature: this fact does not suffice to exclude these constructions from the domain of compounding, rather it enables us to tell that they constitute a *special subtype* of compounding.

The canonical approach, hence, gives us helpful instruments in trying to identify universal tendencies: we expect a certain degree of variation among different languages, which is natural in typological research. In this framework, the common as well as the uncommon *are expected* and the unexpected *is not excluded* from the analysis.

4.2 Classification of Compounds

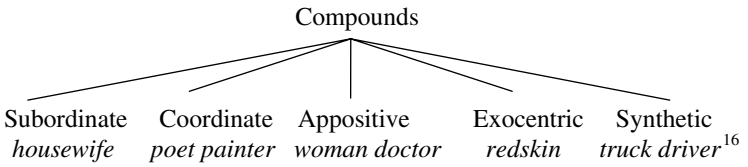
The traditional classifications of compounds that can be found in the literature are inadequate (cf. Bisetto and Scalise 2005 for a detailed discussion). Briefly, most existing classificatory schemes suffer from the serious problem of being based on heterogeneous criteria.

---

[12] The list of properties in (4) is the result of compiling widely accepted proposals which do not hinge on a particular theoretical framework. However, these proposals come from very different frameworks and it may be the case that under closer scrutiny they are shown not to work together.
[13] Cf. Di Sciullo & Williams (1987).
[14] Cf., most recently, Lieber & Scalise (2006) and Booij (2008).

Generalizing, traditional classifications of compounds, whether tacitly or explicitly adopted,[15] share the following basic structure:

(5) *Traditional classifications of compounds:*



Classifications based on such a schema have several drawbacks.

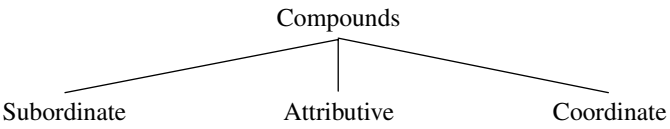
The first problem is that notions such as *subordinate* and *exocentric* are heterogeneous and do not belong to the same level of analysis: *subordinate* refers to the grammatical relationship between the constituents of the compound, while *exocentric* refers to the presence/absence of a head.

Furthermore, a second problem arises when one considers that the classificatory criteria are combinable: e.g. a subordinate compound can be either endocentric (*housewife*) or exocentric (*pickpocket*).

There is a third problem, still: the putative class of *synthetic* compounds is based on a morphological criterion (the head must be a deverbal derivative), which is not applied to any other class: however, synthetic compounds are typically subordinate, and thus the class as a whole is redundant.

Bisetto and Scalise (2005) propose a new classificatory schema, based mainly on the idea that each level of analysis and classification must be consistently based on a single, homogeneous criterion. They propose that the first level be based only on the implicit *grammatical relation* between the constituents. With this criterion, the following schema is obtained:

(6)



[15] Cf. among others, Spencer (1991), Fabb (1998), Olsen (2001), Haspelmath (2002), Booij (2005), etc. Marchand (1969), Bloomfield (1933) and Bauer (2001) proposed instead classificatory schemes which are more elaborated than the one in (5).

[16] Synthetic compounds (also called secondary compounds or verbal-nexus compounds) are formed by a deverbal nominal head, and by a first constituent that is interpreted as the internal argument of the verb underlying the head.

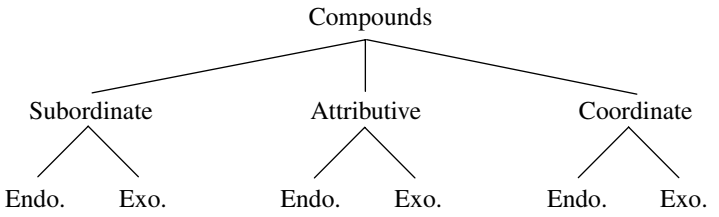
In subordinate compounds there is a grammatical relation of "*complementation*" between the constituents, which is very evident in deverbal compounds (*taxi driver*), where the non-head is interpreted as the internal argument of the verb that underlies the deverbal head. A similar type of relationship can be found also in N+N compounds (also called root compounds, e.g. *apron string*).[17]

In attributive compounds there holds a grammatical relation of *attribution.* Typical examples are A+N structures (*high school*), where the adjectival part expresses a property of the nominal head, or N+A (*ice cold*), where the nominal non-head functions as an adjunct modifying the adjectival head. But N+N structures can also be attributive (*snail mail*),[18] where the non-head noun functions as mere property and is neither referential nor semantically complete (i.e., *snail* functions almost as though it were an adjective).

In coordinate compounds there is a grammatical relation of *coordination*, which is typically a conjunctive coordination (*poet painter)*.[19]

Bisetto and Scalise (2005) propose to introduce a second level of analysis based on the distinction between *endocentric* and *exocentric* compounds, and which is applied homogeneously to the three above-mentioned macro-classes:

(7)



Clear English examples of the six classes of compounds thus individuated are the following, respectively:

[17] While the constituents in a N+N subordinate compound are not linked by an argumental relation, we can still consider this as a subordinative dependency. There is still little knowledge about whether non-deverbal nouns possess some sort of argumental structure. However, some nouns seem to require a further specification by a nominal constituent, or at least show a clear tendency to prefer or favour such a specification: e.g. nouns such as *room* have an intrinsic relational nature and tend to occur with an explicit "complement", cf. *hotel room*, *boiler room*, *conference room*, etc.

[18] In this second case, it would be preferable to speak of an "appositive" relationship, but for the moment being we do not make a further distinction. Besides, V can also be used in an attributive/appositive function, as in the Dutch V+A compound *druipnat* 'drip+wet, soaked' (cf. Guevara & Scalise 2004).

[19] But it can also be a disjunctive coordination. Cf. Mordvin *vest'-kavst* 'once or twice' (Wälchli 2005).

(8)

	Endocentric	Exocentric
SUB	*apron string*	*pickpocket*
ATT	*high school*	*red skin*
CRD	*poet painter*	*mother-child*[20]

This proposal also has consequences for other putative groups of compounds, such as the so-called phrasal and neoclassical compounds, which receive a different classificatory treatment than is usually given to them.

Regarding phrasal compounds such as the following:[21]

(9) *English*: [floor-of-a-birdcage] taste
Dutch: [lach-of-ik-schiet] humor = 'lit. [laugh-or-I-shoot] humour'

the phrasal non-head has a clear attributive function (i.e. it is interpreted as a property, e.g. '*terrible* taste', '*aggressive* humour',), and thus the compound is classified as *attributive*.

In other cases, the phrasal non-head is interpreted as a "complement" of the nominal head, and the whole compound is classified as *subordinate*, cf. (10):

(10) *Dutch*: [oudemannen] huis = 'lit. [old men] house, nursing home'

Regarding neoclassical compounds such as those in (11), consisting of either a free word/stem and a bound stem (or *semiword*[22]), or of two semiwords:

(11) *English*: anthropology [sW+sW]
biochemistry [sW+W]
Italian: odontotecnico [sW+W] = 'lit. tooth technician'
colorificio [W+sW] = 'lit. paint factory'

the relation between the head and the non-head is clearly subordinate (cf. 'the study of man', 'technician of teeth', 'factory of paint').

In conclusion, phrasal and neoclassical compounds need not be classified separately from other "normal" compounds.

4.3 The Notion "Head" and the Position of the Head-Constituent

The presence or absence of a lexical head in a compound can be determined in two different ways, each linked to a distinct notion of morphological head, (i) formal head and (ii) semantic head.

The formal head of a compound is the constituent which shares with – and percolates to – the whole compound all of its formal features: lexical category and subcategorization frame. The whole compound, thus, is expected to have the same distributional properties of its formal head.

---

[20] We intend here *mother-child* as it appears in expressions such as *mother-child relationship*.
[21] Cf. Botha (1984), Lieber (1992).
[22] *Semiword* is the term used in Scalise (1994) to describe these bound stems.

The semantic head of a compound is the constituent which shares with – and percolates to – the whole compound all of its lexical-conceptual information (LCS in short, cf. Jackendoff 1990 and Lieber 2004). The whole compound, thus, is expected to be a hyponym of its semantic head (cf. Scalise, Bisetto and Guevara 2005).

There is not much agreement in the dedicated literature regarding this dichotomy and, as far as we know, there is not a single study that explicitly addresses this issue. Generally, however, there seems to be a tendency to grant the notion of formal head a greater value. In any case, in endocentric compounding the notions of formal and semantic head coincide most of times; this fact can be verified by using the "IS A" test (cf. Allen 1978):

(12) a. *gentil+donna* [A + $N_{[+fem]}$] 'lit. gentle+woman, gentlelady'

=> IS A *donna* => semantic head *donna*
=> IS A $NOUN_{[+fem]}$ => formal head *donna*

b. *capo+stazione* [$N_{[-fem]}$ + $N_{[+fem]}$] 'lit. master+station, station master'

=> IS A *capo* => semantic head *capo*
=> IS A $NOUN_{[-fem]}$ => formal head *capo*

In other cases, however, assigning head status to either constituent is not as easy. Some compounds have more than one suitable candidate, others simply do not seem to have even one (neither for formal head, nor for semantic head):

(13) a. *studente+lavoratore* [$N_{[-fem]}$ + $N_{[-fem]}$] 'student+worker'

=> IS A *studente*
IS A *lavoratore* => semantic head *both?*
=> IS A $NOUN_{[-fem]}$ => formal head *both?*

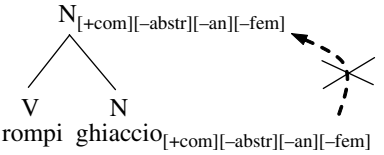
b. *rompi+ghiaccio* [V+$N_{[-fem]}$] 'lit. break+ice, icebreaker'
=> IS NOT A *rompi*
IS NOT A *ghiaccio* => semantic head *none*
=> IS A $NOUN_{[-fem]}$ => formal h. *ghiaccio*?

There are two separate issues here. In the first place, it is not clear what to do with coordinate compounds: some morphologists (cf. Haspelmath 2002, among others) consider that the compounds of the type depicted in (13a) are exocentric; from their point of view, having more than one suitable candidate for the head-role is equivalent to having none. We believe that a descriptively more adequate interpretation of examples such as (13a) is to propose that coordinate endocentric compounds actually have two heads.[23]

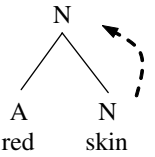
[23] The assumption that there could be two heads in a compound is supported by the semantics of the whole word. On the other hand, formal criteria seem to single out only one candidate as head. Cf. for example It. *nave traghetto* 'ship + ferryboat', where both constituents are plausible semantic heads. From the morphosyntactic point of view, notice that only the left-hand constituent

In the second place, (13b) shows that applying the notion of formal head without taking into consideration the notion of semantic head does not always yield good results: following the formal criterion, one could be tempted to say that in Italian *rompighiaccio* 'icebreaker' is an endocentric right-headed compound (headed by *ghiaccio* 'ice', cf. (14a)), which is clearly nonsense. This is in plain contrast with the widespread idea that possessive compounds in the Germanic languages are actually endocentric, cf. (14b), where the right-hand constituent is said to be also a semantic head by means of a metonymic operation/extension.[24]

(14) a.



b.



skin interpreted metonymically roughly as 'person', 'human being' ([+com][–abstr][+an])

If we only consider the formal notion of head, (14a) and (14b) receive the same representation. Crucially, however, the metonymic extension that can be proposed for the Germanic type (14b) has never been proposed for the Romance type (14a), and we see no compelling reasons to accept it. While metonymy may be a suitable solution for the Germanic exocentric type (14b), it clearly does not offer a way out for the Romance type (14a).

We are not yet able to solve all the problems posed by the notion(s) "head" in compounding, but we can safely conclude that there is no reason to prefer either of the notions "formal head" or "semantic head" over the other. In order to better understand what an exocentric compound is, we must take both of them into consideration.

In what follows, we apply both the notions of "head" seen above in tandem in order to provide with a working definition of endocentric vs. exocentric compounding:[25]

(15) An *endocentric compound* has at least one formal head and at least one semantic head. If an endocentric compound has only one formal head and only one semantic head, then the two must coincide.

(*nave*) bears the same [+feminine] feature that characterizes the whole compound, while the right-hand constituent (*traghetto*) is [-feminine].

[24] Cf. Booij (2005).

[25] Cf. Scalise & Guevara (2006) for an extended treatment of exocentric compounding.

> If a compound has one or more formal heads and no semantic head, it will be considered *exocentric*. If a compound has one or more semantic heads and no formal head, it will also be considered *exocentric*.

In other words, neither the notion of formal head, nor the notion of semantic head, considered separately, suffices to define a compound as endocentric or exocentric.

Furthermore, the identification of the head-constituent in a compound has had a long story. At least the following viewpoints have been argued for:

- in the early '80s a strong hypothesis was put forth, namely, the head in a compound is the **right-hand** constituent (Lieber 1980, Williams 1981);
- in the mid '80s (especially on the basis of the study of Romance languages in the lexicalist framework, cf. Scalise 1983, Corbin 1987) it was established that in some languages the head is the **left-hand** constituent: a consequence of this refinement is that the position of the head could be a good candidate to be a genuine universal parameter;
- but there are other languages still that behave differently; Chinese, for example, has been described as a language with right-headed compound nouns and left-headed compound verbs[26]

We prefer to maintain, in any case, that in every language there is a *canonical position* of the head constituent, which may be disregarded by certain compound-types. This will be especially manifest in the Romance languages, where one cannot really state the position of the head-constituent once and for all.

In the following sections, we analyze compounding data by considering the notions of formal head and the canonical head-position in each studied language.

4.4 Definition of Compound-Type

The analysis in the sections below is based on the data contained in the Morbo/Comp database. The database, as explained above, tries to represent in great detail the peculiarities of compounding in the world's languages. As in any database driven research, merely taking into consideration all the attested entries can suffer from a token frequency bias: the most frequent values in each field (e.g. [N+N] in the field "structure", or N in the field "output category") would outweigh the alternative options, thus blurring the great level of variation among different languages and constituting an insurmountable obstacle for a typological point of view on compounding.

In order to avoid this problem, and drawing from experience in corpus linguistics, we have decided to consider not the tokens (single entries) present in the database but rather the types of compound that can be extrapolated from it.

---

[26] Cf. Packard (2000). This generalization is not without exceptions (cf. Ceccagno & Scalise 2006), but more investigation is needed in order to fully clarify the position of the head in Chinese (cf. Ceccagno & Basciano 2007).

In the present paper, "compound type" is defined as the intersection of the following database fields for each considered language:

(16) Compound-type:

- Output category
- Structure
- Classification
- Position of the head-constituent (formal head)

That is to say, in each language considered, each attested combination of different values for each of the database fields listed in (16) constitutes a different type of compound. In what follows, all the discussed data will be based on this definition of compound-type.

5 Looking for Universals in Morbo/Comp

Before getting into the details of our analysis, some words of caution are in order. This research is currently under development. We are still working on collecting data and the Morbo/Comp database still needs some work of standardization. Our final aim is to apply semi-automatic statistic analyses to the database, extrapolating the typological distribution of the included languages directly from the data. For the time being, our data and our results are still preliminary. However, some clear tendencies can already be seen.

5.1 Methodology

In order to simplify the data we present in this paper, we have divided the database into the following genetically-related groups:[27]

(17) *Romance*: Catalan, French, Italian, Spanish
Germanic: Dutch, English, German, Norwegian, Swedish
Slavic: Bulgarian, Polish, Russian, Serbo-Croatian
East Asian: Chinese (Mandarin), Japanese, Korean

Always for the sake of simplicity, we shall only consider two-member compound-types (however, even compounds with more than two constituents can be analysed as binary formations, cf. Scalise 1983, Guevara 2007).

The first step in the analysis involves selecting the set of characteristic features of compounding that will be taken into consideration. This process of selection is

[27] This grouping is not well-balanced, but it is the best that our data allow us to do at the moment. There is a clear Indo-European bias, and while the languages in the first three groups belong to the same linguistic families, those in the East Asian group neither belong to a single linguistic family nor are typologically homogeneous. We acknowledge these shortcomings.

actually predetermined by the architecture of the Morbo/Comp database and is not likely to change significantly in the near future. Given the "definition" of compound proposed in (3) and here repeated as (18), our analysis will consider the features in (19) for each group's compounds:

(18) [X ℜ Y]Z

(19) a. Z = Output Categories
b. X and Y = Input Categories
c. ℜ = Relation between constituents (Classes)
d. [X Y] = Combination of Categories
e. [X Y] Z = Headedness

In a second step, we calculate the incidence of the different values that each feature in (19) may assume in the compound types of each individual group of languages (cf. 17).

Then, for each feature in (19), we derive the incidence for the whole sample as *the mean incidence* attested for the groups in (17). This mean incidence gives us a clear view on whether there is a preferred value for each feature in the sample (i.e., *cum grano salis*, it shows what the universal or canonical tendency for that feature is).

Finally, we compare the value for *each feature* in *each group* with the mean value for the whole sample, obtaining an indication on the extent to which the language-groups in (17) follow or not the general tendencies extrapolated from the whole sample, and allowing for cross-linguistic comparison.

5.2 Output Categories

We begin by considering the mean incidence of the output category labels for the whole Morbo/Comp sample. The table in (20) shows the five most recurring output categories, which together add up to over 95% of the sample:

(20)

Output cat.	**Mean %**
N	52.24
A	25.85
V	11.63
Adv	6.06
P	0.51

These data tell us that there is a clear scale of preference in compound formation with regard to the lexical category of the output, cf. (21):

(21) **N** > **A** > **V** > **Adv** > **P**

Even though we cannot say *tout simplement* that N is the default output label in compounding (because there is a high incidence of **A**s and **V**s, an incidence that varies from one language to another), we can however safely conclude that **N** occupies, by far, the highest position in compound formation. This is the case also because N is by far the preferred output category in exocentric compounding (e.g. the exocentric Romance patterns V+V, P+N, V+N always give N as output, cf. Section 5.3).

We now compare the mean incidence of the most recurring categories with their distribution across the different language groups. The ranking holds, generally, in all the groups, as it can be seen in (22)

(Legend: RO = Romance, GE = Germanic, EA = East Asian, SL = Slavic):

(22)

Output cat.	**Mean %**	**RO %**	**GE %**	**EA %**	**SL %**
N	52.24	60.29	46.94	47.46	54.29
A	25.85	28.68	26.53	20.34	27.86
V	11.63	5.15	17.01	18.64	5.71
Adv	6.06	5.15	5.44	5.08	8.57
P	0.51		2.04		

Only two possible exceptions to the general tendency appear: first, Romance has (apparently) the same number of verbal and adverbial types; second, Slavic has relatively more adverbial types than verbal types.

It is remarkable that, compared to the general mean, in four cases a significantly higher incidence can be observed: Romance and Slavic show a higher recurrence of N, while Germanic and East Asian show a much higher than expected incidence of V.

5.3 Input Categories

We consider now the input categories in compounding, that is, the category labels of the constituents that take part in a compound-type. The table in (23) shows the five most recurring output categories, which together add up to over 85% of the sample:

(23)

Input cat.	**Mean %**
N	41.01
A	19.76
V	16.13
Adv	5.59
P	2.72

The data suggest the same scale of preference observed for output categories (cf. (21) above), but with a difference at the top of the tables: N is significantly lower as input label than it is as output label:

(24) **N > A > V > Adv > P**

There is, thus, a tendency to produce N in output having categories other than N in input; in other words, exocentric noun-forming compounding has an important impact on the overall primacy of N as output label. There is also a converse tendency: there are more Vs as input categories than as output categories, suggesting that Vs have a tendency to take part in exocentric compounds or as modifiers of endocentric compounds.

We compare now the general mean tendency with the values observed in each group:

(25)

Input cat.	**Mean %**	**RO %**	**GE %**	**EA %**	**SL %**
N	41.01	40.86	37.04	47.50	38.64
A	19.76	20.43	22.90	19.17	16.56
V	16.13	9.32	13.80	31.67	9.74
Adv	5.59	7.17	7.41		7.79
P	2.72	3.23	5.39		2.27

The trend generally holds across the different groups, though showing some slight variation with respect to the individual values. There is however one exception: the East Asian group has a relatively higher incidence of Vs compared to As. Overall, we can observe a clear degree of coherence between the categories attested in Input and in Output in the whole sample.

5.4 Relation Between Constituents (Classification)

We examine now the distribution of the classes of compounds as put forward in Bisetto and Scalise (2005). The mean incidence for each classification type is the following:

(26)

Class.	**Mean %**
SUB	40.06
ATT	32.59
CRD	19.62
OTH	7.73

A clear ranking emerges from the data:

(27) **SUB > ATT > CRD**

The general trend holds in all the groups, though with some variation: Romance and Slavic have a relatively stronger preference for Attributive compounds, while the East Asian group shows a relatively stronger preference for Coordinative compounding (which ranks second for East Asian, in contrast with the general trend). Cf. (28):

(28)

Class.	Mean %	RO %	GE %	EA %	SL %
SUB	40.06	44.85	38.10	37.29	40.00
ATT	32.59	34.56	31.29	28.81	35.71
CRD	19.62	16.18	12.24	32.20	17.86
OTH	7.73	4.41	18.37	1.69	6.43

So in this case the East Asian group does not follow the general pattern, showing a different scale of preference: SUB > CRD > ATT.

In any case, the three classes proposed by Bisetto and Scalise (2005) are largely attested in all the languages in the sample.

5.5 Headedness

Given a two-member compound, the head constituent can be realised logically by one of the following four possibilities: (i) the first constituent is the head, (ii) the second constituent is the head, (iii) both constituents are equally qualified as heads, (iv) no constituent can be said to be the head (i.e. the compound is exocentric). We will signal these four possibilities respectively as: "1", "2", "12", "0". Consider now the mean incidence of these values:

(29)

Headedness	Mean %
2	55.89
0	22.05
1	9.72
12	6.26

We find thus a clear scale of preference for the distribution of heads in compounding:

(30) **Right > No Head > Left > Both**

Overall, there are clearly more endocentric types than exocentric types and a strong preference can be observed for right-headed types (over 55% of all the types). An even more telling figure obtains from the comparison of right headed types against left headed types (55.89% versus 9.72%), confirming the general preference for head-final compounds in the sample.

The general trend holds in all the groups, with a greater degree of variation among the different groups and just one highly significant exception:

(31)

Headedness	Mean %	RO %	GE %	EA %	SL %
2	55.89	36.76	80.95	50.85	55.00
0	22.05	38.24	12.93	22.03	15.00
1	9.72	18.38	2.72	8.47	9.29
12	6.26	5.15	2.04	13.56	4.29

The Romance group shows a different ranking (No Head > Right > Left > Both) with the first two ranks inverted with respect to the general trend. Furthermore, Left headed types are relatively higher than expected.[28]

The Germanic group has a huge majority of right-headed types. Slavic languages follow the trend perfectly. Finally, in the East Asian languages there is a higher than expected number of two-headed types, a fact that is certainly related to this group's preference for coordinative compounding (cf. Section 5.4).

5.6 Combination of Categories

We examine now the attested combinations of categories in the sample. The table in (32) displays the ten most recurring combinations, each accounting at least for ∼2% of all the types. Overall, these ten combinations amount to ∼60% of the whole sample:

(32)

Combinations	Mean %	RO %	GE %	EA %	SL %
[N+N]	18.08	16.91	14.29	25.42	15.71
[A+N]	7.52	8.09	6.80	10.17	5.00
[N+A]	7.39	11.76	8.16	6.78	2.86
[A+A]	6.22	5.88	7.48	5.08	6.43
[V+N]	7.07	4.41	4.08	16.95	2.86
[N+V]	4.40	1.47	4.08	8.47	3.57
[V+V]	4.38	2.94	2.72	11.86	
[Adv+A]	2.31	3.68	2.72		2.86
[Adv+N]	2.27	1.47	4.76		2.86
[A+V]	2.07	0.74	2.72	3.39	1.43

It can be clearly observed that in our sample the privileged structure is **[N+N]**. The remaining combinations have a much lower incidence, and cluster quite closely, making it extremely difficult to draw any conclusions.

[28] This higher than expected presence of left-headed types in Romance (18.38% vs. 9.72% in the whole database) is realized mainly by the prototypical N+N structural combination (summarizing, besides exocentric nouns, right-headed adjectives and some right-headed nouns, Romance has a very strong and populated group of left-headed N+N nouns).

There is also a high degree of cross-linguistic variation and often the general trend is contradicted by a language-specific order of preference. We highlight a few significant cases: Romance shows a relatively stronger incidence of [N+A] and [Adv+A], while East Asian has a relatively higher recurrence of [N+N], [V+N] and [V+V]. Germanic and Slavic follow the general trend quite closely. We interpret the general trend observed in (32) as a three-level scale of preference:

(33) Combinations: **[N+N]** >
[A+N] > [N+A] > [A+A] >
[V+N] > [N+V] > [V+V] > (. . .)

That is to say, on the first level we find the most common type, [N+N], as expected. The second level includes combinations containing Adjectives: an A and an N, or two As; this fact was not to be expected from previous work. Similarly, the third level includes combinations containing Verbs: a V and an N, or combinations of two Vs, a fact that was also unexpected. This three-level hierarchy is confirmed in all the groups considered, with the exception of the 'East Asian' languages, where the third level (compounds including Vs) ranks overall higher than the second level (compounds including As).

We add a last remark to the table in (32): research in compounding has privileged only a minimal portion of the main structures in the languages of the world (mainly [N+N] in all languages, [V+N] in Romance and Chinese, and a few other exceptions). This is even more striking if we examine the table in (34) which lists all the 92 remaining combinations of categories attested in our sample (together adding up to ∼40% of the compound-types in the MorboComp sample used, and still for the most part ignored in the literature).

(34)

Combinations of categories **Remaining structures (92 different combinations) - ordered by incidence**				
[Num+N]	[sN+sA]	[sA+N]	[Adv+Pro]	[Pro+V+extN]
[V+A]	[sA+sN]	[sW+N]	[Adv+V+extA]	[Prt+Adv]
[P+N]	[V+Adv]	[V+N+extN]	[Adv+Vr+N]	[Prt+Pro]
[Adv+Adv]	[N+N+extN]	[A+A+A]	[Conj+Adv]	[sW+A]
[sN+sN]	[P+V]	[Num+V]	[N+Adv]	[sW+sW]
[Adv+V]	[Pro+V]	[P+Adv]	[N+N+A]	[Adv+Ple]
[N+sN]	[Pro+A]	[Ple+N]	[N+N+extA]	[AP+N]
[N+Ple]	[Adv+P]	[CP+N]	[N+N+N]	[PP+V]
[Pro+N]	[NP+N]	[A+PP]	[N+sW]	[Prt+Ple]
[N+PP]	[P+P]	[Adv+Conj+Adv]	[N+V+extA]	[V+Prt]
[sN+N]	[PP+N]	[Adv+PP]	[Ns+Ns]	[VP+N]
[P+A]	[Prt+V]	[N+sA/sN]	[Num+N+extA]	[XP+N]
[Num+A]	[A+N/V]	[N+sA]	[Num+Num]	
[N+Conj+N]	[A+V/N]	[sA+A]	[Num+V+extA]	
[sN+A]	[V+N+extAdv]	[V+Conj+V]	[P+A+N]	
[A+N+extA]	[V+V+extN]	[V+DP]	[P+N+extN]	
[A+A+extN]	[P+N+extA]	[V+Pro]	[Pro+Adv]	
[Adv/P+A]	[A+N+extN]	[[[Neg+V]+N]+extN]	[Pro+Pro]	
[Adv/P+V]	[N+V+extN]	[A+A+N]	[Pro+Prt]	
[Prefixlike+A]	[Num+N+extN]	[Adv+extN]	[Pro+V+extA]	

(Legend: Num = numeral, sW = semiword, sN = semi-noun, sA = semi-adjective, Conj = conjunction, Pro = pronoun, Prt = particle, Ple = participle, Neg = negation, extA = adjective-forming suffix external to the compound, extN = noun-forming suffix external to the compound)

The table in (34) shows that, clearly, there is still a long way to go in order to fully understand compounding structures and their distribution in world's languages.

6 False Universals in Compounding

The search for "true" universals can also reveal that some generalizations that populate the literature and that are part of a sort of a "received wisdom" cannot be held valid when confronted with an adequate set of data. We will briefly comment some such cases.

Generally, what we would call "false universals" have come into being by the overwhelming amount of research on compounding in the Germanic languages, often driving researchers to apply automatically, and hastily, to every language the structural, semantic and pragmatic categories found in the most studied Germanic compound-types.

6.1 Compounds are Only Right-Headed

The widespread idea that compounds are right-headed cannot be maintained. There is extensive data showing that at least some patterns in some languages coherently display left-headedness (cf. Section 5.5).

Usually, the stance that compounds are only right-headed is accompanied by the assumption that compounding is a lexical (morphological) phenomenon, different from the syntax. The rationale thus is as follows: if a putative compound is right-headed, it is a regular morphological structure and it can be formed by the lexical component of the grammar. On the contrary, if a putative compound is left-headed, it would be an irregular morphological structure, and it is better analysed as a syntactic construction of some sort (cf. among many others, Olsen 2001, Spencer 2006), also on the basis that in a left-headed compound often inflectional markings are inserted interrupting the "wordlike string" formed by the two constituents.

The problem is that the arguments are intrinsically hinged on a very restrictive and rigid definition of "compound". Furthermore, the syntactic nature of left-headed compounds is not further pursued, having been swept under the rug of Syntax rather than being better explained. And finally, the insertion of an inflectional morpheme is not quite the same as inserting any kind of lexical material. All these points would deserve specific argumentations, for which we do not have the space here. In sum, the claim that compounds are only right-headed is just a theory-internal construct.

Leaving aside the issue of whether compounding is a lexical or a syntactic phenomenon, a more realistic viewpoint is the following: in general, all languages prefer to form right-headed compounds with a certain extent of language-internal variation (i.e. each compound-type shows a clear and systematic preference). The Romance group has furthermore a strong preference for exocentric compounding. With respect to the most prototypical structural combination, N+N, there is a vast majority of right-headed types, which in some languages is overwhelming (e.g. Germanic languages) while in others it is diminished by the significant and systematic presence of left-headed compounds (e.g. Romance languages, East Asian languages).

6.2 The "Root Compounding Parameter"

In various publications, Snyder (1995, 2001), and also Roeper, Snyder and Hiramatsu (2002) propose the existence of a "Root Compounding Parameter" able to account for some differences (morphological and syntactic) among the world's languages. They argue that a language may accept Complex Predicate Constructions (such as verb-particles, resultative constructions, double objects, etc.) only *if it can productively form N+N compounds*. English, for instance, has the positive value [+] for this parameter, thus permitting complex predicates (besides N+N compounds), while in Spanish and French (and presumably also in Italian) these constructions are un-grammatical because the languages have the [–] value for the parameter. Moreover, English allows also periphrastic constructions with the same semantics of complex predicates, while this is the only possibility in [–] languages: English thus represents a general type of which the Romance languages are a subset with limited characteristics.

Without going into the details of this proposal, we only make one observation: Snyder's parameter implies a strictly binary interpretation of the notion "productivity". We believe that this is incorrect for it is clear that for instance the Romance languages show at least some N+N compounding (especially coordinate and attributive compounds, as well as some subordinate compounds), while they have other very productive compound types (such as V+N A+A, etc.). The relative productivity of compounding as a morphological process is inescapably gradual.

Furthermore, the root compounding parameter is able to account only for right-headed compounding (as a result of leftward movement of modifiers merging with the head of the compound), being thus open to the criticisms expressed in the previous section.

6.3 Head is "Locus Inflectionis"

In compounding, the head is not always the "locus inflectionis" as it has been maintained (e.g. Zwicky 1985, Hudson 1987). We will limit ourselves here only to

pointing out some cases which fall out of this otherwise interesting generalization, undermining its general validity.

Nominal heads in a compound can be mass nouns or event nominals (cf. 35), which as such cannot be pluralized:

(35) a. It. *trasporto latte* 'milk transportation'
b. En. *house cleaning*
cf. **trasport**i** latte*, **house cleaning**s***

To the contrary, although they have no internal head-constituent, subordinate exocentric compounds can be pluralized, as witness the following examples from Italian[29] and English:

(36) a. It. *portacenere* vs. *portaceneri* 'carry-ash = ashtray'
b. En. *pickpocket* vs. *pickpocket**s***

In addition, coordinate compounds can pluralize the constituent in the canonical head-position (37a–b) or both constituents (37c):

(37) a. It. *nave traghetto* *navi traghetto* 'ferry boat'
b. En. *king emperor* *king-emperor**s***
c. It. *studente lavoratore* *studenti lavoratori* 'worker student'

Moreover, there is also evidence showing group-internal and language-internal variation. For instance, Italian attributive N+N compounds mark the plural on the head (38a); the same compound in French marks the plural on both constituents (38b)[30], while in Spanish it can receive plural either on the head or on both constituents (38c):[31]

(38) a. It. *uomo rana* *uomini rana* 'frogman'
b. Fr. *homme grenouille* *hommes grenouille**s***
c. Sp. *hombre rana* *hombres rana* / *hombres rana**s***

Finally, Italian compounds with a 'colour adjective' as head do not pluralize:

(39) a. *giallo limone* 'lemon-yellow'
b. *due maglioni giallo limone* 'two lemon-yellow sweaters '
cf. **due maglioni gialli limone*

Almost all the discussed "false" universals in compounding have to do – in one way or another – with the central notion of "head", its position, its distribution and its

[29] The pluralization of exocentric examples such as (36a) takes place at least in spoken Italian (especially in substandard registers).

[30] Although, as a reviewer rightly points out, plural marks on Fr. *hommes grenouilles* are not realized phonetically, and may be only an artefact of the written language.

[31] However, Spanish shows a strong tendency to prefer head-marking in this type of compound (i.e. only on the left constituent). In any case, double plural marking does not imply ungrammaticality.

behaviour. This shows once again how important this notion is in compounding and how much more research is needed in this direction (hopefully, given the inescapable universality of "head", future research will pay greater attention to cross-linguistic data).

7 Concluding Remarks

Even though our sample is far from being totally representative and balanced, some general tendencies can be observed in compounding in the world's languages. In particular, the following scales of preference in compound formation are attested cross-linguistically:

(40) Output category: **N** > **A** > **V** > **Adv** > (. . .)
Input category: **N** > **A** > **V** > **Adv** > (. . .)
Classification: **SUB** > **ATT** > **CRD**
Headedness: **Right** > **No Head** > **Left** > **Both**
Combinations: **[N+N]** > [A+N] > [N+A] > [A+A] > [V+N] > [N+V] > [V+V] > (. . .)

It is remarkable that the literature has dedicated a great deal of attention to just one case in compounding, instantiated by the combination of the highest values in the rankings in (40), that is: endocentric subordinate right-headed [N+N]N compounds. While this pattern is certainly the canonical instance in compounding in the world's languages, it is by no means the only one. Future work on the typology and on the theory of compounding will necessarily have to shift the tendency shown until now by concentrating on the analysis of the many remaining compound-types.

Similarly, future work in this domain will have to pay more attention to the collection of greater and finer-grained empirical sources. Our work has become possible only after a long period of time devoted to gathering and analysing vast amounts of data. We hope that our instruments will keep growing in the future in order to test the different scales of preference we have presented in this paper against a more representative language sample.

We are aware of the fact that no database can be considered to be definitive. Future developments in the Morbo/Comp database will certainly have consequences on the finer-grained characterization of compounding from a typological point of view in order to grasp more and more interesting insights. Finally, our results require incorporation in a general typological framework for word-formation, and in particular for compounding phenomena, a promising area of inquiry that is still in its initial stages.

References

Ackema, Peter and Ad Neeleman. 2004. *Beyond Morphology: Interface Conditions on Word Formation.* Oxford: Oxford University Press.

Allen, Margaret Reece. 1978. *Morphological Investigations*. Ph.D. Dissertation, University of Connecticut, Storrs.

Baroni, Marco, Emiliano Guevara, and Vito Pirrelli. 2007 (*forthcoming*). Sulla Tipologia dei Composti NN in Italiano: analisi distribuzionale e correlati sperimentali. In *Atti del convegno SLI 2006, Vercelli*, September 21–23 2006. Roma: Bulzoni.

Bauer, Laurie. 1989. Be-heading the word. *Journal of Linguistics* 26:1–31.

Bauer, Laurie. 2001. Compounding. In *Language Typology and Language Universals*, ed. Martin Haspelmath. The Hague: Mouton de Gruyter.

Bisetto, Antonietta and Sergio Scalise. 2005. The classification of compounds. *Lingue e Linguaggio* IV(2): 319–32.

Bloomfield, Leonard. 1933. *Language*. New York: Holt.

Booij, Geert. 2008. Lexical Integrity as a Formal Universal: A Constructionist View. This Volume.

Booij, Geert. 2005. *The Grammar of Words: An Introduction to Linguistic Morphology*. Oxford: Oxford University Press.

Botha, Rudolf. 1984. *Morphological Mechanisms*. Oxford: Pergamon Press.

Ceccagno Antonella and Bianca Basciano. 2007 (*forthcoming*). Neology in Chinese: Classification and Headedness in Chinese Compounds. *Morphology*.

Ceccagno, Antonella, and Sergio Scalise. 2006. Composti del cinese: analisi delle strutture e identificazione della testa. In *La Cina e l'altro*, ed. Annamaria Palermo. Napoli: Il Torcoliere.

Chomsky, Noam. 2005. Three Factors in Language Design. *Linguistic Inquiry* 36: 1–22.

Comrie, Bernard. 2003. On explaining Language Universals. In *The New Psychology of Language,* Vol. II, 1195–209 ed. Michael Tomasello. Vol: II, I. Mahwah, N.J.: Lawrence Erlbaum.

Corbett, Greville G. 2005. The canonical approach in typology. In *Linguistic Diversity and Language Theories*, 25–49 eds. Zygmunt Frajzyngier, Adam Hodges and David S. Rood. Amsterdam: Benjamins.

Corbett, Greville G. 2008. Universals and Features. This Volume.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen : Max Niemeyer Verlag.

Costello, Fintan J., and Mark T. Keane. 2000. Efficient creativity: Constraint-guided conceptual combination. *Cognitive Science* 24: 299–349.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William. 2004. Logical and typological arguments for Radical Construction Grammar. In *Construction Grammar(s): Cognitive and Cross-language Dimensions*, eds. Mirjam Fried and Jan-Ola Östman. Constructional Approaches to Language, vol. I. Amsterdam: John Benjamins.

Di Sciullo, Anna-Maria. 2005. *Asymmetry in Morphology.* Cambridge (MA): MIT Press.

Di Sciullo, Anna-Maria, and Edwin Williams. 1987. *On the Definition of Word*. Cambridge (MA): MIT Press.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language*, 68 (1): 81–138.

Fabb, Nigel. 1998. Compounding. In *Handbook of Morphology*, 66–83 eds. Andrew Spencer and Arnold M. Zwicky. Oxford: Blackwell.

Fitch, W. Tecumseh., Marc D. Hauser, and Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implications. *Cognition* 97: 179–210.

Gagne, Christina L., and Edward J. Shoben. 1997. Influence of thematic relations on the comprehension of modifier–noun combinations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23: 71–87.

Gottfried, Gail M. 1997. Comprehending compounds: evidence for metaphoric skills? *Journal of Child Language* 24: 163–186.

Guevara, Emiliano. 2007. Binary Branching and Linguistic Theory: Morphological Arguments. In *Proceedings of the "XXXII Incontro di Grammatica Generativa", Firenze, 2–4 March 2006*, 93–106 eds. M. Cecilia Picchi, and Alan Pona. Alessandria: Edizioni dell'Orso.

Guevara, Emiliano, and Sergio Scalise. 2004. V-Compounding in Dutch and Italian. *Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset*, XI. Madrid: I.U. Ortega y Gasset.

Haspelmath, Martin. 2002. *Understanding Morphology*. London: Arnold.

Haspelmath, Martin. 2007. Pre-established Categories don't Exist: Consequences for Language Description and Typology. *Linguistic Typology* 11: 119–132.

Haspelmath, Martin, Matthew Dryer, David Gil, and Bernard Comrie, eds. 2005. *The World Atlas of Language Structures*. (Book with interactive CD-ROM) Oxford: Oxford University Press.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch (2002). The faculty of language: What is it, who has it, and how does it evolve? *Science* 298: 1569–1579.

Hudson, Richard A. 1987. Zwicky on Heads. *Journal of Linguistics* 23: 109–32.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge (MA): MIT Press.

Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. Cambridge (MA). MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Jackendoff , Ray. 2007 (to appear). Compounding in the Parallel Architecture and Conceptual semantics. Ms.

Jackendoff, Ray, and Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97: 211–225.

Lees, Robert B. 1960. *The Grammar of English Nominalizations*. The Hague: Mouton.

Levi, Judith N. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

Lieber, Rochelle. 1980. *On the Organization of the Lexicon*. Ph.D Dissertation. MIT (distributed by IULC 1981).

Lieber, Rochelle. 1992. *Deconstructing Morphology*. Chicago: University of Chicago Press.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

Lieber Rochelle, and Sergio Scalise. 2006. The Lexical Integrity Hypothesis in a New Theoretical Universe. *Lingue e Linguaggio* V.I: 7–32.

Mairal, Ricardo, and Juana Gil, eds. 2006. *Linguistic Universals*. Cambridge: Cambridge University Press.

Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word-Formation: A Synchronic-Diachronic Approach.* München: C.H. Beck. 2nd edition.

Olsen, Susan. 2000. Composition. Chapter 87 in *Morphology. A Handbook of Inflection and Word Formation*, 897–916 eds. Geert Booij, Christian Lehmann, and Joachim Mugdan. Berlin: Walter de Gruyter.

Olsen, Susan. 2001. Copulative Compounds. A Closer Look at the Interface Between Morphology and Syntax. In *Yearbook of Morphology 2000*, eds. Geert Booij, and Jaap van Marle. 279–320.

Packard, Jerome. 2000. *The Morphology of Chinese*. Cambridge: Cambridge University Press.

Plank, Frans, and Elena Filimonova. 1999. *The Universals Archive*. Universität Konstanz. http://typo.uni-konstanz.de/archive/

Pinker, Steven, and Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95: 201–236.

Roeper, Thomas, William Snyder, and Kazuko Hiramatsu. 2002. Learnability in a Minimalist Framework: Root Compounds, Merger, and the Syntax-Morphology Interface. In *The Process of Language Acquisition*, ed. Ingeborg Lasser. Frankfurt: Peter Lang.

Scalise, Sergio. 1983. *Generative Morphology.* Dordrecht: Foris.

Scalise, Sergio. 1994. Morfologia. Bologna: Il Mulino.

Scalise, Sergio, Antonietta Bisetto, and Emiliano Guevara. 2005. Selection in Compounding. In *Morphology and its Demarcations*, eds. Wolfgang U. Dressler et al. Amsterdam: Benjamins, 133–50.

Scalise, Sergio, and Emiliano Guevara. 2006. Exocentric Compounding in a Typological Framework. *Lingue e Linguaggio* V.II: 185–206.

Snyder, William. 1995. *Language Acquisition and Language Variation: The Role of Morphology*. Ph.D Dissertation, MIT.

Snyder, William. 2001. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77: 324–342.

Sohn, Ho-Min. 1999. *The Korean Language*. Cambridge: Cambridge University Press.

Spencer, Andrew. 1991. *Morphological Theory*. Oxford: Blackwell.

Spencer, Andrew. 2006. Morphological Universals. In *Linguistic Universals*, 101–129 eds. Ricardo Mairal, and Juana Gil. Cambridge: Cambridge University Press.

Talmy, Leonard. 2003. Concept Structuring Systems in Language. In *The New Psychology of Language*, Vol. II, 15–46 ed. Michael Tomasello. Mahwah, N.J.: Lawrence Erlbaum.

Tomasello, Michael. 2003. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge (MA): Harvard University Press.

Tomasello, Michael. 2006. Acquiring linguistic constructions. In Handbook of Child Psychology, Volume 2, "Cognition, Perception, and Language", 255–298, eds. William Damon, Richard M. Lerner, Deanna Kuhn, Robert S. Siegler. New York: Wiley.

Wälchli, Bernhard. 2005. *Co-compounds and Natural Coordination*. Oxford: Oxford University Press.

Williams, Edwin. 1981. On the notions 'lexically related' and 'head of a word'. *Linguistic Inquiry* 12: 245–74.

Wisniewski, Edward J. 1997. When concepts combine. *Psychonomic Bulletin & Review* 4: 167–183.

Wisniewski, Edward J., and Dedre Gentner. 1991. On the combinatorial semantics of noun pairs: Minor and major adjustments to meaning. In *Understanding word and sentence*, 241–284 ed. Greg B. Simpson. Amsterdam: North-Holland.

Zwicky, Arnold M. 1985. Heads. *Journal of Linguistics* 22: 1–29.

Universals and Features[1]

Greville G. Corbett

Abstract Greenberg's paper on universals (1963) contains an interesting set of generalizations relating to features. It is a good time to review the issues involved in establishing universals of features. These verge on the philosophical at one extreme, while at the other they concern the practical question of how we present and gloss examples. Various initiatives concerned with standardization, taken broadly, are under way, and it is important that they should be fully informed by the linguistic issues. There are two main areas to discuss: the *Analysis problem* and the *Correspondence problem*.

The Analysis problem: for a given language, we need to be able to justify the postulation of any feature (such as number or case). Equally, for each feature in the language we need to be able to justify the set of values postulated (for example: singular, dual, paucal and plural; nominative, accusative and genitive). For some languages the analysis is trivially simple, in others it is exceptionally complex (for some there have been long-running debates). In this context, it is worth reviewing the work of the Set-theoretical School, given its undoubted relevance for typology. The difficulties posed by hybrids will be discussed; this leads naturally to typological hierarchies and the 'Canonical' approach in modern typology.

The Correspondence problem: as typologists we need to be able to justify treating features and their values as comparable across languages. This is not straightforward, and yet a good deal of typology, including enterprises such as the *World Atlas of Language Structures*, depends upon it. The problem has a second, more subtle version. Even within a single language, features and their values do not necessarily line up consistently. In Bayso, the number system of nouns and verbs interact in a complex way. In Romanian, the genders of nouns and adjectives differ, and there are many more such examples. Here a typological perspective can inform the analysis of a single language and, of course, a typology which ignored these languages would be considerably impoverished.

G.G. Corbett
Surrey Morphology Group

[1]The support of the ESRC under grants RES-051-27-0122 and RES-062-23-0696, and of the AHRC under grant B/RG/AN4375/APN18306 is gratefully acknowledged. I thank Anna Kibort and two anonymous referees for helpful comments.

S. Scalise et al. (eds.), *Universals of Language Today*, 129–143,
DOI: 10.1007/978-1-4020-8825-4_7,

Features are an area where the concerns of the typologist meet those of computational linguists, formal linguists, fieldworkers, in fact linguists in many different guises. As we put increasing theoretical weight on features, it is important to review our assumptions and check our progress in understanding them.

Keywords Features · universals · typology · morphosyntax · standardization

Greenberg (1963) includes several interesting generalizations on features. Universals 30–45 are relevant – in particular to morphosyntactic features. Since that time there has been substantial typological research into features. Equally in formal syntax, features have taken on an ever increasing significance. Indeed, Miller and Sag (1997: 579) call feature structures 'the fundamental construct used to model linguistic entities'. It is now an appropriate time to review the issues involved. Some are profound, and will always be the subject of debate, while others are highly practical, concerning standardization and the presentation and glossing of examples. As is becoming generally accepted, we shall use the term 'feature' for gender, number, and so on, and 'value' for feminine, neuter ... and singular, dual, plural ... (compare Ramat 1998, Corbett 2006b). We shall concentrate on morphosyntactic features, taking the term strictly to imply features that are relevant to morphology and syntax. We are not here concerned with purely morphological features (for which see Corbett and Baerman 2007). We also distinguish morphosyntactic features from morphosemantic features, which are not relevant to syntax: see Stump (2005) and Corbett (forthcoming) for discussion; an example of a morphosemantic feature would be tense in the numerous languages where tense is morphologically distinguished on the verb but where this has no impact on syntax.

1 Why Features?

It is worth reminding ourselves why we use features. There are different motivations which have converged, so that features are now shared across a very wide range of linguistic work, from the most theoretical to the highly applied.

1.1 An Abbreviatory Device

In one sense, features come 'free', since they do not increase the expressive power of a grammar. We can interpret a symbol like NP_{pl} as a single ornate symbol (Halle 1969, Gazdar, Klein, Pullum, and Sag 1985: 20–21). At this level of analysis, features are a useful abbreviatory device.

1.2 A Way of Making Generalizations

The other side of the coin is that features allow us to make generalizations. They allow us to say, for example, that within a given language the same distinctions of

number occur across different constructions (agreement within the noun phrase as opposed to within the clause) and yet are realized differently across lexemes (thus *this : these :: runs : run*).

1.3 The Basis for Typology

Having isolated the distinctions which we model using features, it is natural to typologize across them, as in Greenberg (1963). As with all typology, we need to consider carefully whether we are comparing like with like, an issue to which we return in Section 6.

2 Usefulness of Features and Issues of Standardization

Features are central to various initiatives concerned with standardization, taken broadly, some of which are currently under way. It is important that such initiatives should be fully informed by the linguistic issues.

2.1 EAGLES

The report on morphosyntactic annotation (Leech and Wilson 1996) is an early attempt to grapple with the issues. It was restricted to languages of the European Union, and does not fully distinguish part of speech and semantic subcategories from morphosyntax. Tags suggested for particular languages were included rather than their being rigorously compared with the general set established for a wider range of languages.

2.2 Lexical Markup Framework (LMF)

The International Organization for Standardization (ISO), in particular Technical Committee ISO/TC 37, *Terminology and other language resources*, Subcommittee SC 4, *Language resource management,* is working on ISO 24613 'Language resource management – Lexical markup framework'. (Revision 14 was circulated in mid July 2007.) The goals, as stated in the introduction (p. 5) are as follows:

> Lexical Markup Framework (LMF) is an abstract metamodel that provides a common, standardized framework for the construction of computational lexicons. LMF ensures the encoding of linguistic information in a way that enables reusability in different applications and for different tasks. LMF provides a common, shared representation of lexical objects, including morphological, syntactic, and semantic aspects.
>
> The goals of LMF are to provide a common model for the creation and use of electronic lexical resources ranging from small to large in scale, to manage the exchange of data between and among these resources, and to facilitate the merging of large numbers of different individual electronic resources to form extensive global electronic resources. The ultimate goal of LMF is to create a modular structure that will facilitate true content interoperability across all aspects of electronic lexical resources.

A previous draft was quite disappointing from a linguist's standpoint; however, the Committee has taken on board comments from linguists, and the latest draft is considerably improved.

2.3 E-MELD (Electronic Metastructure for Endangered Languages Data) and GOLD (General Ontology for Linguistic Description)

E-MELD had two primary objectives: contributing to preserving data on endangered languages, and helping to develop the infrastructure for effective collaboration between electronic archives (Aristar Dry 2002). The first objective was focused on best practice, in a variety of areas. So far as it concerned morphosyntactic markup, the direction was not so much to suggest a standard, as to ensure that non-significant differences in annotation should not hamper further understanding and analysis. This was consonant with the second objective, and led to initial work on an ontology of linguistic concepts (Farrar and Langendoen 2003). Work continues in this direction, taking account of the notion of canonicity (discussed in Section 5.2).

2.4 The Leipzig Glossing Rules

The Leipzig Glossing Rules (Comrie, Haspelmath and Bickel 2004), which build on earlier work particularly by Lehmann (1983), represent a bottom-up approach to standardization. They are available at:

http://www.eva.mpg.de/lingua/files/morpheme.html

At the simplest level it is eminently sensible that we should use the same symbols (e.g. '=' for clitic boundary) and the same abbreviations. We should be certain whether a colleague wishes to indicate perfect tense or perfective aspect.

While the proposals in the Leipzig Glossing Rules may seem low-level and relatively uncontroversial, this perspective on them may change quite rapidly if one tries to apply them consistently for a large and diverse set of examples (as in Corbett 2006a). The first observation is that glossing to the level of detail required is a demanding undertaking. And second, quite substantial issues come to the surface surprisingly quickly. The reason is that there are numerous problems with the analysis of features (which will come to shortly) and these come to the fore in glossing. Before considering those further, let us ask what is, or at least might be, universal.

3 Can There be Universals in this Area?

Given the genuine difficulties of analysis, it is worth asking what we may hope to identify as universal. A simple suggestion can be found in Zwicky (1986: 988),

namely that 'universal grammar should permit only a finite number of attributes and values – indeed [...] universal grammar should provide finite lists of the attributes and values available for service in a particular grammar'. Zwicky points out the difficult with the approach, as put to him by Gerald Gazdar:[2]

> Gazdar's challenge (in personal communication) cuts deeper. He observes that there is a serious correspondence problem involved in talking about 'the illative case' in two different languages: what allows us to identify the two grammatical cases? Similarly for other agreement properties, other head properties, and foot properties as well.
>
> This is not the place to mount a full response, but I believe it is possible to require that every property on the lists have semantic concomitants. I am not maintaining here that these properties are to be IDENTIFIED with semantic features; grammatical categories are virtually always arbitrarily distributed (from the semantic point of view) in the lexicon to some extent. I am suggesting that a head or foot property is never a FULLY arbitrary and language-particular categorization of words and phrases: it has a semantic core that runs across languages.
>
> Zwicky (1986: 988–989)

Zwicky's suggestion, then, is that morphosyntactic features always have a semantic core, and it is this core which allows comparison. We return to this issue in Section 6 below.

4 The Analysis Problem: Features

For a given language, we need to be able to justify the postulation of any feature. Since as we have seen features are an abbreviatory device, we have to ask persistently whether each is needed. This is essential for the typologist, since there is the danger of always finding the features we expect, especially if we take functions as the starting point. An interesting example of an argument that an accepted feature is not actually required is Spencer's (forthcoming) analysis of Hungarian, in which he argues that there is no need to recognize a case feature for that language.

4.1 Phonological Form

We might reasonably assume that in order to postulate a morphosyntactic feature, and its various values, we would require that for each there would be an inflected form (unique in its phonology) which could be explained only in terms of the particular feature and value. The discussion is usually for justifying particular values, but it is necessary at the feature level too. The existence of a unique form may seem an obvious requirement, but in fact it is too strong. There are situations where a feature is justifiable even though there is no dedicated form to support it. Thus Chumakina,

[2] Zwicky (1986: 988) also reports a challenge by Geoffrey Pullum, suggesting that a fixed list is not plausible, given the remarkable variety which is already known. Like Zwicky, I am not deterred by this.

Kibort and Corbett (2007) argue that the feature person is required in the grammar of Archi, on the basis that it is required if the resolution rules of the language are to be stated simply and in ways which are plausible in cross-linguistic terms.

4.2 Conditions Versus Features

It is important for typologists to distinguish clearly between morphosyntactic features and conditions. As a brief example, consider these data on agreement with conjoined noun phrases in Russian:

(1) Predicate agreement with conjoined noun phrases in Russian

subject type word order	animate		inanimate	
	N	%PL	N	%PL
subject-predicate	115	100	67	85
predicate-subject	89	84	114	28

The data are from a corpus of literary works (Corbett 1983: 106, 128, 130). They indicate clearly that the word order has a major influence on the agreement form selected. The plural, the semantically agreeing form, is more likely to be found than the singular if the subject precedes the predicate. We would not add word order to the list of morphosyntactic features. Rather we say that it is a possible condition of the use of a morphosyntactic feature (number in this instance). The data show the effect of a second condition, namely animacy. If the subject is semantically animate, plural agreement is more likely than if it is inanimate. (Russian does have animacy as a subgender, but it is rather semantic animacy which influences the agreement choice we are considering.) For valid typological comparisons we must distinguish between morphosyntactic features and conditions on their use. Thus respect is often a condition on the use of a feature (often number, sometimes person) and yet it may also be a feature in its own right, with a dedicated form (Section 4.1). Conditions have interesting properties (for instance, they have consistent effects cross-linguistically); for discussion and key examples see Corbett (2006a: 176–205).

5 The Analysis Problem: Values

Equally, for each feature in the language we need to be able to justify the set of values postulated (for example: singular, dual, paucal and plural; nominative, accusative and genitive). For some languages the analysis is simple, in others it is exceptionally complex (as demonstrated by discussions in the literature that have persisted over decades).

5.1 Set-Theoretical Approaches

In this context, it is worth reviewing the work of the Set-theoretical School, given its undoubted relevance for typology (and the fact that fifty years have just passed since the first meeting of the famous seminar on mathematical linguistics in Moscow). The famous mathematician Andrej Kolmogorov posed the following questions (van Helden 1993: 138):

> "What exactly do we mean when we say that two words are in the same case?"
> "How many cases does the Russian language possess?"

There was a flowering of interesting work on such questions, including particularly relevant work by Zaliznjak (1973). It is carefully surveyed by van Helden (1993) and a good introduction is Meyer (1994). In brief, Zaliznjak and others worked out careful and consistent methods for determining the feature and value inventory of a language (and a good deal of substandard work in typology could have been avoided if their legacy were better known). Typically the expected features and values are established, but less clear instances often emerge too, that is, the formal approach highlights interesting data and challenges. A phenomenon recognized and documented within this approach is the instances of non-autonomous case values (Zaliznjak 1973: 69–74). Here there is no form uniquely associated with a particular value, but the value is justified on the basis of syncretic forms. There are comparable non-autonomous gender values (Corbett 1991: 150–154); an alternative term is *genus alternans*; see Igartua (2006) and references there for discussion of the development of such instances in Indo-European.

Set-theoretical accounts are 'brittle', by which I mean that finding a single additional lexeme or context may be sufficient to invalidate an analysis. This is exactly what is required in terms of falsifiability – it is clear what constitutes a counterexample. And yet morphosyntactic feature systems often seem to be less rigid than such analyses allow. We consider an instance of this in the next section.

5.2 Hybrids

A significant problem for set-theoretical approaches is hybrids, that is, controllers whose feature specification varies according to the target. A familiar example is *committee* nouns in various varieties of English.

(2) *Committee* nouns in spoken American English and British English (Levin 2001: 109)

	verb		relative pronoun		personal pronoun	
	N	% plural	N	% plural	N	% plural
LSAC	524	9	43	74	239	94
BNC	2086	32	277	58	607	72

Note: LSAC = Longman Spoken American Corpus; BNC = British National Corpus (section on spoken language).

We see that the number value for *committee* and similar nouns varies according to the target. It is not straightforwardly singular or plural. Early researchers in the Set-theoretical School were aware of the issue. One obvious approach was to treat hybrids as having a different feature value. This would work if all hybrids behaved alike. However, this turns out not to be the case. Evidence demonstrating this for gender is reported in Corbett (1991: 183–184), and for number in Corbett (2006a: 213). Since each hybrid can be different, each would require a different feature value, and the number of values would be hugely extended.

The alternative is to restrict the number of feature values, essentially to those required by non-hybrid nouns (already the notion of 'canonicity is coming into play, to be discussed further below). The problem of hybrids is then dealt by two interrelated means. First by typological hierarchies; in the example in question, this would be the Agreement Hierarchy:

(3) Agreement Hierarchy (Corbett 2006a: 207)
attributive > predicate > relative pronoun > personal pronoun

This hierarchy allows us to constrain possible agreement patterns as follows:

> For any controller that permits alternative agreements, as we move rightwards along the Agreement Hierarchy, the likelihood of agreement with greater semantic justification will increase monotonically.

Thus the variability in the morphosyntactic feature specification of hybrids is constrained, rather than varying freely. The second part of the analysis is conditions (as discussed in Section 4.2); these may involve semantic information, down to the detail of particular lexical items.

This general approach is consonant with the Canonical approach in modern typology. Here we set up clear definitions and take them to the logical end point, defining a theoretical space before asking where particular examples fit into it. In the problem just discussed, our definitions (specifically our definitions of feature values) will be based on controllers which take consistent agreements. Hybrids are then non-canonical. Furthermore, particular feature values in a given language may be more or less canonical. While in this way we avoid the explosion in the number of feature values, we leave open the possibility that a particular feature might have some certain values and further values of less certain status (a classic instance is the Russian case system; see, among others, Zaliznjak 1973, Comrie 1986, Mel'čuk 1986/2006, Corbett forthcoming).

6 The Correspondence Problem: Cross-Linguistic

As typologists we need to be able to justify treating features and their values as comparable across languages. This is not straightforward, and yet a good deal of typology, including enterprises such as the *World Atlas of Language Structures*, depends upon it. At the level of features, provided we are concentrating on morphosyntactic features, there is rarely a problem. That is, we know whether we are

comparing case across languages, as opposed to gender or person. At the level of values, however, which is as Gazdar stated the correspondence problem, things are more difficult. There are two ways forward. The first is to avoid the problem by lowering our sights to comparing systems only in terms of size (that is, the number of distinct values). We can make generalizations about the maximal and minimal systems (as in Section 8.3 below). And, as Greenberg did (1963), we can make claims about interactions between feature values; however, even this requires us to be able to compare at least some feature values cross-linguistically.

The second way forward, the one Zwicky suggests, is that 'it is possible to require that every property on the lists have semantic concomitants' (1986: 988). At the level of features, this is plausible. Thus gender always has a semantic core (Aksenov 1984, Corbett 1991: 8–69) and we could extend that to every morphosyntactic feature. There is the issue of case, which Zwicky (1992) treated as indirect, that is, as not 'associated directly with prototypical, or default, semantics' (1992: 378). Nevertheless, the argument for a correspondence with semantics, taken broadly, can be made. This is also the stance of Svenonius (2007). However, Gazdar's point concerns values, and here the issue is more difficult.[3] For some values, cross-linguistic comparison is straightforward: feminine gender is the value which includes nouns denoting females, and the interesting typological considerations are what other nouns may be included in this gender value, how the assignment rules overlap or are distinguished from others, and whether or not the feminine is the default gender value. We need to define the core meanings and functions: we call a gender value the feminine if it includes nouns denoting females, whether or not it also includes diminutives. Similarly we call a case value the dative if used for recipients, whether or not it can also be governed by prepositions. But it does not follow that all values can be compared in this way. While gender features always have a semantic core, it is not clear that all gender values have a semantic core. The issue needs to be resolved first at the level of the individual languages. If it proves to be the case that some values have no semantic core, then we should compare features first in terms of the semantic core, and then in terms of the possible remaining values which fall outside that core.

7 The Correspondence Problem: Intra-Linguistic

While Gazdar stated the correspondence problem in cross-linguistic terms, there is an analogous, more subtle intra-linguistic version. Even within a single language, features and their values do not necessarily line up consistently.

[3] This is a point where we have to be explicit about whether we are discussing features or values; discussion often flits between the two, and the intention is sometimes not clear.

An instance of a feature which does not correspond within a given language is number, of the nominal and verbal types. Nominal number is concerned with the number of entities; it may appear on targets by agreement. Thus in *Mary runs*, number is of the nominal type, reflecting the fact that Mary is one individual, and not that there is a single running event. Verbal number indicates the number of events, or the number of participants in events. The two work rather differently, as documented in Corbett (2000: 243–264). However, they can appear together, as illustrated by Georgian (Corbett 2000: 254–255). Only nominal number is a morphosyntactic feature.

Turning to the intra-linguistic lack of correspondence of values, a well-studied instance is the gender system of Romanian. Here nouns are assigned to three genders, while agreeing targets distinguish only two. In other words, there are three controller genders and two target genders (Corbett 1991: 150–154). There are various other examples, and we take a less familiar one for illustration, namely the number system in the Cushitic language Bayso (Hayward 1979, Corbett and Hayward 1987, Corbett 2006a: 172–174). In Bayso, the number systems of nouns and verbs interact in a complex way. Nouns mark four numbers (general, singular, paucal and plural), while verbs show singular agreement (and gender agreement) for general and singular, plural agreement with the paucal, and masculine singular agreement with the plural.

We should note that both for Romanian and for Bayso we are not considering small numbers of irregular lexical items. We are looking at the normal system which involves substantial proportions of the lexicon. This section shows that for some languages there may be no straightforward response to the questions 'how many gender values?' and 'how many number values?' Here a typological perspective can inform the analysis of individual languages and, of course, a typology which ignored these languages would be considerably impoverished.

8 What is Universal?

Given the care that must be taken over issues of correspondence, we may wonder what we can hope for when looking for universals of features. The strategy advocated here is to opt for the simplest outcome, and give that up only if it can be demonstrated to be unattainable.

8.1 The Simplest Possibility

We should start from the simplest possibility, which would be a Zwicky-type list. If again we restrict ourselves to clearly morphosyntactic features, it is clear that the well-established agreement features (phi-features) all qualify. These are gender, number and person. In addition, case is clearly relevant to syntax.

There are two further, less obvious morphosyntactic features. Respect is often conveyed by the use of other features, thus it is often a condition on the use

of number or person (Section 4.2). However, it may also, if rarely, appear as a morphosyntactic feature.[4] This is shown by Muna (an Austronesian language spoken on Muna, off the southeast coast of Sulawesi).

(4) Number and politeness markers in Muna (van den Berg 1989: 51, 82)

'go' (second person)	singular	plural
neutral	o-kala	o-kala-amu
polite	to-kala	to-kala-amu

These equivalents of '(you) go' vary according to number and politeness; *to-* marks polite address, irrespective of number.

Definiteness too can occur as a morphosyntactic feature, as these Norwegian data show:

Norwegian (Bokmål, Torodd Kinn and Tore Nesset, personal communications)

(5) det ny-e hus-et mitt
DEF.N.SG new.DEF.N.SG house(N)-DEF.N.SG my.N.SG
'my new house'

(6) mitt ny-e hus
my.N.SG new-DEF.N.SG house(N)[INDEF]
'my new house'

Clearly definiteness marking is sensitive to the syntactic environment, and so appears to quality as a morphosyntactic feature, in a small number of languages. For more on definiteness marking in Scandinavian languages see Delsing (1993: 113–184), and for recent discussion see Hankamer and Mikkelsen (2002).

8.2 A Possible Need to Extend: Kayardild

The possibility of maintaining a relatively small list of morphosyntactic features (in the strict sense, excluding morphosemantic features) is severely challenged by data from Kayardild. As well as ordinary cases, Kayardild has various verbalizing cases (Evans 2003). Consider the verbalizing dative (V_DAT), which is used for beneficiaries. Its marker, which repeats through the noun phrase, is *-maru-*. The surprising thing is that this marker takes regular verbal inflections:

[4] For interesting discussion of the development of the plural pronoun as a politeness marker in Icelandic, and comparison with various other languages, see Guðmundsson (1972). However, in Icelandic, only pronoun choice is involved: the original dual pronouns, which became plurals, and the plural pronouns, which became polite pronouns, took plural verb agreement. The predicate adjective may be singular or plural for polite plural pronouns (Comrie 1975: 409, citing Einarsson 1945: 134), but again we are dealing with values of number, and not a distinct morphosyntactic feature.

Kayardild (Evans 2003: 215)

(7) ngada waa-jarra wangarr-ina ngijin-maru-tharra thabuju-maru-tharra
1SG.NOM sing-PST song-MOD_ABL my-V_DAT-PST brother-V_DAT-PST
'I sang a song for my brother.'

In example (7) we see tense marked on different elements of the noun phrase; Evans gives comparable examples for aspect, mood and polarity. Thus while these features are often morphosemantic (since they need not be referred to by rules of syntax), this is not evident in Kayardild. We may analyse such examples in different ways; see Evans (2003) and Corbett (2006a: 138–140) for discussion. If one believes that tense, aspect, mood and polarity are features of the clause, then marking of these features on more than one item is symmetrical marking, and hence not (canonical) agreement. (In dependency approaches this would be agreement.) Whatever our analysis, the Kayardild data show that we cannot limit the list of morphosyntactic features to the obvious core instances without careful argumentation.

8.3 Minimal and Maximal Systems?

It may not be possible to achieve typologies which specify the possible configurations of features (for instance, we might have liked to claim that a language cannot have a paucal unless it has a dual, but Bayso appears to be a counter-example). However, we may be able to specify the smallest and largest systems.

For most morphosyntactic features, the **smallest system** is the logically possible smallest system, that with two values. Thus many languages have two genders only (see Corbett 2005). Similarly number systems with just two values are commonplace.

One feature that might seem problematic here is person. Greenberg's universal number 42 (1963: 113) states that: 'All languages have pronominal categories involving at least three person and two numbers.' However, we would not necessary treat pronominal distinctions as morphosyntactic, and so we might find two-valued systems without conflict with Greenberg's universal. And indeed, it has been claimed that the Daghestanian language Archi makes a binary morphosyntactic distinction between first and second person on the one hand and the third person on the other (Chumakina, Kibort and Corbett 2007). There are distinct pronominal forms, but no morphosyntactic evidence to split first and second persons. The relevant contrasts are shown in this paradigm:

(8) Person agreement in Archi

PERSON	NUMBER	
	SG	PL
1	gender agreement	Ø-
2	gender agreement	Ø-
3	gender agreement	gender agreement

In the plural, the marker for first and second persons is Ø- (the bare stem is used). As mentioned above (Section 4.1), there is no dedicated form for person agreement, since this marker is also that for genders III and IV in the third person. Yet we need a morphosyntactic person feature both to account for forms in the paradigm, and for the resolution rules. Other small systems are discussed in Cysouw (2003: 127–139); note that Cysouw discusses individual paradigms, so that a paradigm with only two person values may apply only to some items and not necessarily be indicative of the features values available in the language as a whole.

When we come to specify the **largest system**, this is naturally harder. See, for instance, the discussion of large case systems in Comrie and Polinsky (1998). However, for number, considerable progress has been made. It seems that the largest systems of number contain five values (see Corbett 2000: 39–42). The interesting point is that they appear in different configurations. Thus Mele-Fila has the values: singular, dual, paucal, plural, greater plural; while Sursurunga has singular, dual, paucal, greater paucal, plural.[5]

9 Conclusion

Features are an area where the concerns of the typologist meet those of computational linguists, formal linguists, fieldworkers, in fact linguists in many different guises. As we put increasing theoretical weight on features, it is important to review our assumptions and check our progress in understanding them. A reasonable strategy is to try for the simplest typology: fixed lists of features, of values, and of configurations of values. The latter two lists are different, since we know that the largest systems do not necessarily include all the attested values. Of course, our lists are open to challenge from every new piece of research, but we should be able to construct them with sufficient plausibility for us to wish to scrutinize claims for necessary extensions with a degree of scepticism.

References

Aksenov, A. T. 1984. K probleme èkstralingvističeskoj motivacii grammatičeskoj kategorii roda. *Voprosy jazykoznanija* no. 1.14–25.

Aristar Dry, Helen. 2002. E-MELD: Overview and Update. International Workshop on Resources and Tools in Field Linguistics, Las Palmas 26–27 May 2002. [available at: http://linguistlist.org/emeld/documents/index.cfm]

Berg, René van den. 1989. *A Grammar of the Muna Language* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 139). Dordrecht: Foris.

Chumakina, Marina, Anna Kibort, and Greville G. Corbett. 2007. Determining a language's feature inventory: person in Archi. In *Endangered Languages* (special issue of *Linguistische Berichte,* number 14), ed. Peter K. Austin and Andrew Simpson, 143–172. Hamburg: Helmut Buske.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51: 406–418.

[5] Corbett (2000: 29) is correct here for Sursurunga; there is a typo on page 42 of that publication.

Comrie, Bernard. 1986. On delimiting cases. In *Case in Slavic*, ed. Richard D. Brecht and James Levine, 86–106. Columbus, OH: Slavica.

Comrie, Bernard, Martin Haspelmath, and Balthasar Bickel. 2004. *The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme glosses*. [available at: http://www.eva.mpg.de/lingua/files/morpheme.html

Comrie, Bernard and Polinsky, Maria. 1998. The great Daghestanian case hoax. In *Case, Typology and Grammar: In Honor of Barry J. Blake* (Typological Studies in Language 38), ed. Anna Siewierska and Jae Jung Song, 95–114. Amsterdam: John Benjamins

Corbett, Greville G. 1983. *Hierarchies, Targets And Controllers: Agreement Patterns in Slavic.* London: Croom Helm.

Corbett, Greville G. 1991. *Gender.* Cambridge: Cambridge University Press.

Corbett, Greville G. 2000. *Number.* Cambridge: Cambridge University Press.

Corbett, Greville G. 2005. The number of genders (chapter and map). In *The World Atlas of Language Structures*, ed. Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie, 126–129. Oxford: Oxford University Press.

Corbett, Greville G. 2006a. *Agreement.* Cambridge: Cambridge University Press.

Corbett, Greville G. 2006b. Linguistic features. In *Encyclopedia of Language and Linguistics, 2nd Edition, Volume 7*, ed. Keith Brown, 193–194. Oxford: Elsevier.

Corbett, Greville G. (forthcoming). Determining morphosyntactic feature values: the case of case. To appear in *Case and grammatical relations: papers in honour of Bernard Comrie*, ed. Greville G. Corbett & Michael Noonan. Amsterdam: John Benjamins.

Corbett, Greville G. and Matthew Baerman. 2007. Prolegomena to a typology of morphological features. *Morphology* 16: 231–246.

Corbett, Greville G. and Richard J. Hayward. 1987. Gender and number in Bayso. *Lingua* 73.1–28.

Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

Delsing, Lars-Olof. 1993. *The internal structure of noun phrases in the Scandinavian languages: a comparative study.* Lund: Department of Scandinavian Languages, University of Lund.

Einarsson, Stefán. 1945. *Icelandic: grammar, texts, glossary*. Baltimore: Johns Hopkins Press.

Evans, Nicholas. 2003. Typologies of agreement: some problems from Kayardild. In *Agreement: a typological perspective* (Special issue of *Transactions of the Philological Society* 101, no. 2), ed. Dunstan Brown, Greville G. Corbett and Carole Tiberius, 203–234. Oxford: Blackwell.

Farrar, Scott and D. Terence Langendoen. 2003. A Linguistic Ontology for the Semantic Web. *GLOT International* 7, no. 3: 97–100.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. G*eneralized Phrase Structure Grammar.* Blackwell: Oxford.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of Language,* ed. Joseph H. Greenberg, 73–113. Cambridge, MA: MIT Press. [Paperback edition published 1966; page references to this edition.]

Guðmundsson, Helgi. 1972. *The Pronominal Dual in Icelandic* (University of Iceland Publications in Linguistics 2). Reykjavík: Institute of Nordic Linguistics.

Halle, Morris. 1969. How not to measure length of lexical representations and other matters. *Journal of Linguistics* 5: 305–308.

Hankamer, Jorge and Line Mikkelsen. 2002. A morphological analysis of definite nouns in Danish. *Journal of Germanic Linguistics* 14: 137–175.

Hayward, Richard J. 1979. Bayso revisited: some preliminary linguistic observations – II. *Bulletin of the School of Oriental and African Studies, University of London* 42: 101–32.

van Helden, W. Andries. 1993. *Case and gender: Concept formation between morphology and syntax (II volumes)* (Studies in Slavic and General Linguistics 20). Amsterdam: Rodopi.

Igartua, Iván. 2006. *Genus alternans* in Indo-European. *Indogermanische Forschungen* 111: 56–70.

Leech, Geoffrey and Andrew Wilson (main authors). 1996. *EAGLES: Recommendations for the Morphosyntactic Annotation of Corpora (EAGLES Document EAG–TCWG–MAC/R).* [available at: www.ilc.cnr.it/EAGLES96/pub/eagles/corpora/annotate.ps.gz

Lehmann, Christian. 1983. Directions for interlinear morphemic translations. *Folia Linguistica* 16: 193–224.

Levin, Magnus. 2001. *Agreement with Collective Nouns in English.* Almqvist & Wiksell.

Mel'čuk, Igor. (1986). Toward a definition of case. In *Case in Slavic*. ed. Richard D. Brecht and James Levine, 35–85. Columbus, OH: Slavica, [revised version in: Igor Mel'čuk. 2006. *Aspects of the Theory of Morphology,* ed. David Beck. Berlin: Mouton de Gruyter, 110–179.]

Meyer, Peter. 1994. Grammatical categories and the methodology of linguistics: Review article on van Helden, W. Andries: 1993, '*Case and gender . . . '. Russian Linguistics* 18: 341–377.

Miller, Philip H. and Ivan A. Sag. 1997. French clitic movement without clitics or movement. *Natural Language and Linguistic Theory* 15: 573–639.

Ramat, Paolo. 1998. On categories and categorizations. In *Language Change and Typological Variation: In Honor of Winfred P. Lehmann on the Occasion of His 83rd Birthday: II: Grammatical Universals and Typology* [=Journal of Indo-European Studies, Monograph 31], ed. Carol F. Justus and Edgar C. Polomé, 330–347. Washington DC: Institute for the Study of Man.

Spencer, Andrew. (forthcoming). Does Hungarian have a case system? To appear in *Case and grammatical relations: papers in honour of Bernard Comrie*, ed. Greville G. Corbett & Michael Noonan. Amsterdam: John Benjamins.

Stump, Gregory T. 2005. Word-formation and inflectional morphology. In *Handbook of Word-formation* (Studies in Natural Language and Linguistic Theory 64), ed. Pavol Štekauer and Rochelle Lieber, 49–71. Dordrecht: Springer.

Svenonius, Peter. 2007. Interpreting uninterpretable features. *Linguistic Analysis* 33: 375–413.

Zaliznjak, Andrej A. 1973. O ponimanii termina 'padež' v lingvističeskix opisanijax. In *Problemy grammatičeskogo modelirovanija*, ed. Andrej A. Zaliznjak, 53–87. Moscow: Nauka.

Zwicky, Arnold M. (1986). German adjective agreement in GPSG. *Linguistics* 24: 957–990.

Methods for Finding Language Universals in Syntax

William Croft

Abstract There are two broad theoretical traditions in research on syntactic universals, the typological and the generative. While the theoretical differences between the two approaches are well known, less known but equally important is the fact that the methods used to find and justify syntactic universals are very different between the two approaches. The generative approach, illustrated here with Baker's (2003) analysis of verbal vs. nonverbal predication, proceeds "one language at a time", and draws on whatever constructions are deemed relevant to support the universal underlying structure. However, the "one language at a time" method often produces universals that fail a wider crosslinguistic comparison; but the choice of constructions used to counter anomalous evidence is an instance of methodological opportunism (Croft 2001). A rigorous method of syntactic argumentation is proposed, which is close to best practice in typological research.

Keywords Syntactic universals · typology · generative grammar · syntactic argumentation · methodology

1 Typological (Functional) and Generative (Formal) Approaches to Syntactic Universals

Research on syntactic universals is currently carried out in two theoretical traditions in linguistics, the typological (or Greenbergian) and the generative (or Chomskyan) approaches*. These two approaches to language universals have been practiced for

W. Croft
University of New Mexico

*This chapter is a revised version of my closing statement from a dialogue course on Syntactic Categories cotaught with Mark Baker at the Summer Institute of the Linguistic Society of America, Cambridge, Mass., in July 2005. I express my thanks to the audience at the LSA Institute as well as the audience at the Language Universals conference in Bologna for their questions and comments. I especially thank Mark Baker for the intellectual debate he engaged in, as well as two anonymous referees for Springer Verlag. Needless to say, none of the above, especially not Mark, are responsible for any misunderstandings or other errors found herein.

S. Scalise et al. (eds.), *Universals of Language Today*, 145–164,
DOI: 10.1007/978-1-4020-8825-4_8,

forty years now (in fact, the occasion of this publication is the fortieth anniversary of the publication of Greenberg 1966a). The differences in the two approaches to language universals have therefore been discussed at some length (see for example Comrie 1981/1989, 1984, 2003; Coopmans 1983, 1984; Hawkins 1985, 1988; Croft 1995, 2001; Newmeyer 1998, 2005).

As a consequence, the major differences in the two approaches are taken to be well known. Universals in typology are inductively derived from a crosslinguistic sample of grammatical structures, often in the form of an implicational relationship. Explanations for language universals are usually functional in character, that is, in terms of how linguistic functions are encoded in linguistic form, in the context of communication and cognitive storage and processing. Because of this mode of explanation, the typological approach is sometimes called the functional-typological approach, and more recently has been subsumed under the more general category of functionalist approaches. (However, many functionalists focus their attention largely on semantics and discourse function, and so have said relatively little about syntax or syntactic universals.)

In contrast, universals in generative grammar are deductively derived from hypotheses about innate language capacity (the so-called poverty of the stimulus argument) and grammatical patterns in a particular language or languages. Explanations for language universals are usually formal in character, that is, formal structures are posited as universal structures in an innate Universal Grammar. Some other syntactic theories reject the notion of an innate Universal Grammar, or remain agnostic about it, but still propose universals that are formal in character. These latter theories are often grouped with generative grammar under the general label of formalist approaches to language.

It might be thought that this is all to be said on the topic, and indeed with a few exceptions (such as Newmeyer and myself), the generative and typological approaches, or more generally the formal and functional approaches, have gone their own ways since the debates published in the 1980s. In this chapter, I argue that such a judgement is premature. An important difference, perhaps the most important difference, between the two approaches has not been illuminated by those authors comparing the two approaches: namely, HOW each approach identifies universals of syntax (but see Croft 2001, 2004, 2005, 2007). The examination of the different methods sheds considerable light on why the universals found in each approach, and the explanations offered for them, differ. While I will also argue that the typological approach is methodologically more sound, I hope that this chapter will stimulate generative linguists as well as typologists to examine more carefully their methodological commitments (in the sense of Laudan 1977) and prompt further dialogue between the two approaches that may lead to a better understanding of their differences.

In conversations with linguists and occasionally in print over the years, I have encountered four views about the relationship between the formal and typological (or more generally functional-typological) approaches to syntactic universals. Most if not all of the four views have been expressed by both generative linguists and

typologists, not to mention other linguists who try not to "take sides" in the debate. The four views are summarized below. I use the more general terms 'formal' and 'functional', since there is some basis in the belief that the views presented below apply to the more general approaches to grammar.

(1) Formal and functional approaches are basically complementary. Formalists and functionalists are examining different aspects of language, and they are theoretically compatible. This is the most conciliatory view, raising the possibility that the two approaches can be integrated into a single theory of language.
(2) Formal and functional approaches are in competition, but formal approaches are said to "go deeper into the grammar" than functional approaches. That is, formalists and functionalists are examining the same sorts of phenomena in language, but they differ in that formal approaches offer "deeper" analyses in some sense (what this means will be discussed further below). Although this view generally assumes that the formal approach is superior, some linguists take the view that the allegedly "superficial" universals found in the typological approach may be complementary, or at least merit some sort of "deeper" explanation.
(3) Formal and functional approaches are essentially contradictory. That is, formalists and functionalists are examining the same sorts of phenomena in language, but their ways of doing so are not reconcilable. This view represents the two as theories of the same scientific phenomenon that are in competition, such that one is better than the other, at least until a third approach superior to both comes along.
(4) Formal and functional approaches are incommensurable. That is, formalists and functionalists are doing totally different things. In this view, the two approaches are incompatible, and cannot even be compared. This last view is the most pessimistic, but appears to be a conclusion sometimes drawn from the near absence of dialog between the two approaches over the past two decades or so.

I believe that two of these views can be fairly easily ruled out. The formalist and functionalist approaches are neither incommensurable nor complementary. Both approaches take as their object of scientific study the investigation of grammatical form. Thus it is possible to compare how each approach attacks the problem of analyzing grammatical form—i.e., they are not incommensurable (view 4). However, nor are they complementary. It is true that functional explanations for typological universals are usually formulated in terms of how linguistic function is encoded in grammatical form. But a functional linguist is still examining grammatical form, and an explanation in terms of function is in competition with an explanation in terms of form—i.e. the two approaches are not complementary (view 1). (Some formalist and some functionalist research is indeed complementary. Formal syntactic analysis is at least in principle independent of meaning or function. Conversely, many functionalists. e.g. cognitive semanticists and discourse analysts, study only

meaning and function. These two areas are therefore complementary. However, they are unlikely to be susceptible to integration.)

This conclusion leaves us with two of the views outlined above. Formalists and functionalists are examining the same phenomena. Both are positing grammatical categories and grammatical distinctions supported by linguistic behavior. Both use a range of distributional facts to support their analyses of grammar. Nevertheless the result—the language universals proposed—is very different. The generative, and more broadly formalist, universals of syntax are universals of structural properties of linguistic form that constrain possible formal grammars (the "principles" of more recent generative grammar; other formal theories such as Head-driven Phrase Structure Grammar and Lexical Functional Grammar have similar principles). The typological universals of syntax are universals that constrain how functions are expressed in linguistic form.

Why is this so? In the next section, I argue that it is because the methods of syntactic analysis are fundamentally different in the two approaches.

2 Methodological Differences Between Generative and Typological Approaches to Syntactic Universals

In Section 1, I stated that the methodological differences in how generative linguists and typological linguists look for syntactic universals have been overlooked in the literature describing the differences between the two approaches. (I return to the narrower description of these approaches because many functionalists do not take a crosslinguistic approach either.) This is not exactly true. There is one methodological difference: typologists look at a broad range of languages; indeed, that is what most people think of when they think of typology. Nevertheless, this is not quite the right description of the difference between the generative and typological approaches. After all, there are generative studies of grammatical phenomena in many languages, and there are some generative linguists, notably Mark Baker, who bring in a diverse set of languages simultaneously in order to analyze a particular grammatical phenomenon.

The real methodological difference has to do with how a generative linguist or a typologist approaches the problem of examining the same grammatical phenomenon in more than one language. The generative method can be described as "one language at a time". This method is in fact a continuation of the structuralist method, under the structuralist principle that all of the grammatical structures in a language "hang together" (*tout se tient* in the structuralist dictum). In the generative method, the linguist examines some grammatical phenomenon, essentially a construction or family of constructions, and then looks for other grammatical phenomena (constructions) that appear to interact with the phenomenon under examination. The various combinations of constructions will produce a distributional pattern of grammatical and ungrammatical sentences, and a theory will be proposed to explain the pattern.

To give a simple textbook example (a more complex example will be given below), we can compare the following English sentences:

(5) The boys ate all the pizza.
(6) The road extends ten miles into the mountains.
(7) The stranger gave the kids brownies.
(8) The stranger gave brownies to the kids.

Sentences (5)–(8), like all sentences, represent a combination of different grammatical phenomena or constructions. The constructions we are specifically interested in are the ones involved in distinguishing the arguments of a predicate, by position relative to the verb and by whether the argument is governed by a preposition or not, and the active voice. We can extend our analysis by showing how the argument structure interacts with another construction, the passive voice:

(9) All the pizza was eaten by the boys.
(10) *Ten miles is extended by the road into the mountains.
(11) The kids were given brownies by the stranger.
(12) *Brownies were given the kids by the stranger.
(13) Brownies were given to the kids by the stranger.

The interaction with the passive yields a mismatch in distribution, namely that sentences (10) and (12) are ungrammatical in (American) English. The theory that has been proposed to explain this pattern, in its most general formulation, is to posit an underlying category of Direct Object (however that category is to be characterized); to define the Passive (however it is defined) as applying only to Direct Objects; and not assigning Direct Object status to *ten miles* in (6) or to *brownies* in (7). These three proposals will together explain the ungrammaticality of (10) and (12): the relevant arguments are not Direct Objects, so the Passive cannot be formed.

This sort of argumentation is what is meant by "going deeper" into the grammar of a language. A linguist examining only sentences (5)–(8) might assume that *ten miles* in (6) and *brownies* in (7) must be analyzed just like *the pizza* in (5), *the kids* in (7) and *brownies* in (8). But looking at another construction, the passive, shows that this assumption is incorrect. And looking at still more constructions may show that the argument structures of these and other verbs will differ in still other ways, requiring explanation. Thus, the analysis involves looking at a large number of constructions in a language.

For crosslinguistic studies, this process is repeated. Starting from a comparable phenomenon, e.g. the form of arguments of different predicates in the active voice, a generative linguist will look for other grammatical phenomena that interact with the first phenomenon, and then try to explain the distributional patterns found in that interaction. The constructions examined in the other languages may not be the same constructions examined in the language one started with; for example, the language may lack a passive, or the language may have case inflections on nouns where English does not. Ideally, however, the underlying theoretical constructs will be the same, or at least some of them will be the same. Those which are the same will be the syntactic universals that have been uncovered. Since the underlying theoretical

constructs will manifest themselves in each language, examination of a single language can reveal syntactic universals of the type that generative linguists propose.

The typological approach to language universals uses a different method for crosslinguistic comparison. In the typological method, a typological linguist begins by looking across languages, preferably a large sample selected according to some sampling algorithm. The typologist then looks at the same construction or the same set of related constructions across those languages. For example, one can examine the form of arguments in the active voice, or the relationship between the active and passive voices. Constructions are defined as the "same" or comparable in terms of functional equivalence. For example, predicates and their arguments are treated as the same across languages if they are semantic equivalents, e.g. they mean 'eat' or 'give' or 'extend', and the semantic role is 'thing eaten' or 'thing given' or 'measure of extent' respectively (see Section 4 below). In the typological approach, unlike the generative approach, it is not assumed that one can find syntactic universals in one language, even though each language expresses the same meanings or functions that are used as the basis for crosslinguistic comparison. This is because the syntactic universals are actually universals of how meaning or function is encoded in grammatical form, and the encoding can vary in ways that are not predictable from how meaning is encoded in one particular language.

In the next section, I critically examine one recent generative analysis leading to a syntactic universal, the category of Pred and its behavior in Baker's theory of syntactic categories (Baker 2003), and compare it to a typological analysis of some of the same grammatical phenomena.

3 Critique of Structuralist/Generative Syntactic Argumentation

In this section, I offer a typologist's perspective on generative syntactic argumentation leading to the positing of generative syntactic universals. This perspective will end up being critical. Yet at least as important is the fact that it is possible. That is, we can compare the generative and typological approach to syntactic universals, and in so doing, bring empirical evidence to bear.

The generative analysis I choose is Baker's hypothesis that Nouns and Adjectives as parts of speech require a syntactic node with a Pred functional category in predication, while Verbs do not (Baker 2003: 35). This is one of the major defining features of Verbs as opposed to non-Verbs in Baker's universal theory of parts of speech. The specific universal is:

(14) Nouns and Adjectives require Pred in predication; Verbs do not.[1]

Baker's analysis is chosen partly as a representative example of generative argumentation for syntactic universals. But it is mainly chosen because Baker takes more of a crosslinguistic comparative perspective than other generative linguists, who may argue for syntactic universals on the basis of a single language; and because

[1] Baker later suggests that Pred is in fact 'conflated' in Verbs, that is, Verbs are the product of a conflation of Pred and Adjective (Baker 2003: 87–88).

Baker is unafraid to present problematic language examples for his theory and to propose theoretical explanations for the anomalies.

Baker motivates his universal Pred for Nouns and Adjectives by noting that 'nouns and adjectives in many languages need help in order to be main clause predicates; they must appear in construction with a copular verb like *be*' as in English *Chris is hungry/a skier* (Baker 2003: 31). Baker continues that 'the frequent need for a copular element to appear with predicate adjectives and nouns but not verbs is a reflection of the fact that. . .nouns and adjectives are never predicates in and of themselves; they can only count as predicates in a derivative sense, by being part of a more articulated structure' (ibid.).

This structure is the Pred node in the syntactic configuration. However, despite motivating the existence of Pred by the requirement of a copula for predicate nominals and adjectives in languages like English, Baker actually argues that many copulas, including English *be*, are not manifestations of Pred. Baker arrives at this conclusion by comparing the interaction of the presence/absence of *be* with two other constructions, secondary predicates (untensed small clauses in his terms), and participial verb forms (Baker 2003: 40):

(15) The poisoned food made Chris (*be) sick/an invalid.
(16) Chris *(is) dying.

The absence of *be* in secondary predicate nominals and adjectives, and the presence of *be* in participial verb forms, indicates the lack of a strict correlation of Pred with Noun/Adjective and not with Verb.

However, Baker argues that the overt copula is indeed a manifestation of Pred in two other languages, Edo and Chichewa. Baker provides the following arguments for Edo: the copula *yé* is required in secondary predicate nominals and adjectives (see e.g. (17)); a floating-quantifier-like form is positioned before *yé* parallel to its position before a Verb, rather than before the Adjective/Noun (and after *yé*; (18)); and *yé* cannot be nominalized, undergo predicate cleft, or partake in serial verb constructions (Baker 2003: 41–43):[2]

(17) *oȼ ∼ yá e ȼ!gógó wo ȼ∼ro ȼ∼
it made bell long
'It made the bell long.' (Baker 2003: 41, ex. 46b)

(18) ozó (tòbo ȼ!rè) yé (*tòbo ȼ!rè) mòsèmòsè
Ozo (by.self) PRED (by.self) beautiful
'Ozo alone is beautiful.' (Baker 2003: 43, ex. 51)

In his argument for *yé* as Pred in Edo, Baker thus invokes interaction with two other constructions, secondary predication and the position of floating-quantifier-like elements. But if these constructions are reliable indicators of Pred, then the following complex biconditional universal should be valid:[3]

---

[2] Abbreviations in examples: 3SGM third singular masculine, COP copula, IND indicative, NOM nominative, NPST nonpast, NSF noun suffix, NsS neuter singular subject, PST past.

[3] Example (19) is shorthand for a set of interlocking biconditional universals.

(19) Copula occurs in nominal/adjectival secondary predication ≡ Floated quantifier occurs in same position with respect to Copula as it does with respect to Verb ≡ Copula is not nominalized ≡ Copula is not clefted ≡ Copula does not occur in serial verb constructions.

This biconditional universal is implied by Baker's theoretical analysis of the relationship between the underlying theoretical construct Pred and other underlying theoretical constructs involved in the generative analysis of secondary predication, floating quantifiers, nominalization, clefting and serial verbs. From a typologist's point of view, Baker has derived this biconditional universal from the examination of just two languages, English (which has the value False for all the propositions in (19)) and Edo (which has the value True for all the propositions in (19)). In effect, Baker is predicting that all languages are either exactly like English (all values False) or exactly like Edo (all values True) with respect to the grammatical properties in (19). In typological terms, this is a very strong claim to make on the basis of a tiny sample; and in fact, Baker has to revise (19) in light of the third language he examines, Chichewa.

For Chichewa, Baker observes that the Copula *ndì* is not verblike in its morphology, and cannot form a causative, and concludes that it is a manifestation of Pred. However, in a footnote Baker notes that *ndì* also does not occur in secondary predication (Baker 2003: 45fn11). Baker accounts for this anomalous behavior by arguing that the Chichewa secondary predication uses the subjunctive, and is therefore not tenseless; only the presence/absence of an overt copula with a tenseless secondary predication counts as evidence of overt Pred.

The facts of Chichewa combined with the facts of Edo (and English) require a reformulation of the complex biconditional universal in (19), adding a biconditional about causatives, and revising the first proposition to exclude subjunctive secondary predications (revisions to (19) are in italics):

(20) Copula occurs in *tenseless* nominal/adjectival secondary predication ≡ Floated quantifier occurs in same position with respect to Copula as it does with respect to Verb ≡ Copula is not nominalized ≡ Copula is not clefted ≡ Copula does not occur in serial verb constructions ≡ *Copula cannot form a morphological causative*.

The revised biconditional universal in (20) covers English, Edo and Chichewa. Nevertheless, the fact that the universal in (19) had to be revised to (20) with the addition of just one more language gives the typologist little faith that the revised universal in (20) will survive in that form when even just one further language is added. Instead, it appears that the syntactic universal in (14), which cannot be directly tested empirically, is simply assumed to be correct; and when the empirically testable universal in (19) based on the predictions of (14) plus the rest of generative theory is problematic, then only the empirically testable universal is revised.

Baker then turns his attention to languages without overt Pred, specifically Mohawk, Hebrew and Quechua. In Mohawk, Baker examines the interaction of noun incorporation with Pred. For Baker, noun incorporation is theoretically

correlated with unaccusative predications, so he predicts it will be unacceptable to incorporate arguments into predicate nouns and adjectives, since nouns and adjectives by themselves lack Pred and therefore cannot assign a theta role to the incorporated argument. (This is a brief illustration of the way in which the syntactic universal (14) combines with other parts of generative theory to produce an empirically testable hypothesis.) However, in Mohawk, predicate adjectives do allow incorporation, although predicate nominals do not, as predicted:

(21) ka-nuhs-íyo
NsS-house-be.good
'That house is good.' (Baker 2003: 71, ex. 101c)

(22) *ka-'nerohkw-a-núhs-a'
NsS-box-Ø-house-NSF
'(That) box is a house.' (Baker 2003: 71, ex 103b)

Baker argues that Mohawk Adjectives are really Verbs: they inflect for tense and person agreement, allow morphological causatives, and allow floated quantifier like elements. (Baker 2003: 70, 71fn23).

In effect, Baker is predicting that there is an interaction of (null) PRED with noun incorporation, such that absence of PRED implies that incorporation cannot take place. But this is falsified for Mohawk "Adjectives", so a further claim is introduced, such that if incorporation takes place in the absence of PRED, then the lexical items in question inflect for tense/person, allow morphological causatives, etc. Baker's analysis/argument can be formulated as a complex implicational universal:

(23) Null PRED $\supset$ (Incorporation $\supset$ (Presence of tense/person inflection & Morphological causative & Occurrence of floated quantifier elements))

In Hebrew, Baker appeals to a different construction, external possession (also known as possessor ascension). This construction is also hypothesized to be an indicator of unaccusativity, which incidentally implies the biconditional universal in (24):

(24) Incorporation $\equiv$ External Possession

External possession, like incorporation, is predicted to be ungrammatical with predicate nominals and adjectivals; and this is indeed the case in Hebrew:

(25) Ha-simla hata lebana (*li)
the-dress be.PST white (to.me)
'The (*my) dress was white.' (Baker 2003: 73, ex. 106a)

(26) Ha-nahaq-et hayta rofa (*le-Rina)
the-driver be.PST doctor (to-Rina)
'The (*Rina's) driver used to be a doctor.' (Baker 2003: 73, ex. 106c)

However, recall that the empirically testable universal must follow from the combination of Pred with other theoretical assumptions of generative grammar. Here Baker must actually alter the other theoretical assumptions of generative grammar to make the empirical prediction follow from the introduction of Pred into the theory,

specifically that the dative expression is inside the lexical projection and to substitute m-command for c-command (Baker 2003: 73, 74).

Finally, Baker examines Quechua. Baker argues that if a language has a morphological causative, it will be restricted to Verbs (Baker 2003: 53–54); this grammatical property has already been invoked for Chichewa and Mohawk. However, Quechua provides a counterexample: there are morphological causatives for nouns and verbs. Baker proposes that if Nouns/Adjectives can take a morphological causative, then it must be distinct in form from the Verbal causative (Baker 2003: 55). However, Imbabura Quechua provides a counterexample to that hypothesis, so Baker proposes that if Nouns/Adjectives can take a morphological causative, and it is not distinct in form from the Verbal causative, then it must be a case of accidental homophony (Baker 2003: 57). These qualifications in order to accommodate Quechua in Baker's syntactic universal (14) require the hypothesis of the complex implicational universal in (27):

(27) Absence of PRED ⊃ (Morphological causative for N/A ⊃ (Nondistinct morphological causative for N/A ⊃ Accidental homophony))

These examples will have to suffice to illustrate the generative approach to syntactic universals. A syntactic universal referring to theoretical constructs such as Pred is proposed to account for certain patterns in the predication of verbs, nouns and adjectives; this is the universal in (14). The data from various languages, taken one at a time, is argued to support the universal in (14) by invoking other constructions, and also other principles of generative theory that account for the interaction of the other constructions with the original construction under study (in this case, predication of different parts of speech). The consequence of this methodology is that hypothesized universals must be added to the universal in (14) that reflect the predictions of (14) combined with other principles of generative theory. These latter universals can be directly compared to the universals proposed in the typological approach, and empirically evaluated. However, these latter universals are revised and added to as each language is added to the generative analysis using the "one language at a time" method. The original universal, formulated only in terms of theoretical constructs, is left unchanged.

From a typologists's perspective, there are two serious problems with the generative methodology for deriving syntactic universals and evaluating them with respect to evidence from crosslinguistic comparison. The first has to do with the "one language at a time" method. In this method, the languages are examined individually before they are compared, or more precisely, one starts with one language, and successively compares a second language to the first, then a third to the second, and so on, modifying the hypothesis as one goes on. This method contrasts with the typological method, in which one examines a broad sample of languages to begin with, and formulates hypotheses based on the evidence from the broad sample as a whole.

The problem with the "one language at a time" approach is that there is a high likelihood that a lot of effort will be expended on proposed interactions between constructions that will in fact not hold universally. An example is the occurrence

of the copula. As we noted above, Baker argues that the presence/absence of an overt copula is not to be equated with the presence/absence of Pred; for example, the English copula *be* is not a manifestation of Pred. Why does the English copula occur, then? Baker argues that *be* 'appears when the lexical head of the clause cannot bear finite tense and agreement morphology' (Baker 2003: 40). This explanation for the presence of a copula is in fact structuralist in origin (Stassen 1997: 65), but it fits with the other characteristic of the generative method, namely invoking another construction to account for the grammatical behavior of the construction one begins with. In this case, one begins with the presence/absence of the copula, and the other construction or grammatical phenomenon that is invoked is the occurrence of tense and agreement morphology. Stassen calls this hypothesis the Dummy Hypothesis (Stassen 1997: 65–76):

(28) Zero copula occurs when verbs lack inflection or when the inflection is zero (e.g. in 3SG or present tense); nonzero inflections require a "dummy" copula.

The Dummy Hypothesis is a biconditional universal: Copula ≡ Nonzero inflections. Its empirical predictions can be described by the tetrachoric table in (29):

(29)

	zero copula	full copula
unmarked TAM	+	−
marked TAM	−	+

Stassen uses a sample of 410 languages and demonstrates that there are a large number of languages that violate both predictions of the Dummy Hypothesis. There are many languages with no zero inflections but a zero copula, such as Sinhalese (Gair 1970: 144, 145, 45, cited in Stassen 1997: 68):

(30) mahattea e-nəw-a
gentleman come-NPST-IND
'The boss comes/will come.'

(31) mahattea a-aw-a
gentleman come-PST-IND
'The boss came.'

(32) unnæhee hungak prəsiddə kene-k
3SG.M very famous person-NOM
'He is/was a very famous person.'

And there are many languages with (at least some) zero inflections but always an overt copula, such as Cambodian (Jacob 1968: 69, 202, cited in Stassen 1997: 74):

(33) vì:ə tʏu phsa:r
he go market
'He goes/went/will go to market.'

(34) mən-s nùh cì:ə kru:
man that COP teacher
'That man is a teacher.'

Another large-scale typological study of a generative syntactic universal is Gilligan's (1987) study of a generative universal, the pro-drop parameter. The pro-drop parameter posited a relationship between four constructions, illustrated for Spanish (which possesses them all) and English (which lacks them all) in (35)–(38):

(35) *Null thematic subjects*
Hemos trabajado todo el día.
'*Worked all day.'

(36) *Null nonthematic subjects*
Llueve.
'*Is raining.'

(37) *Subject inversion*
Salió María.
'*Left Mary.'

(38) that-*trace violation*
Quién dijiste que salió temprano?
'*Who did you say that __ left early?'

These four binary grammatical properties (grammaticality vs. ungrammaticality of the four constructions) entail sixteen possible language types. Of the sixteen, three types are predicted to occur by the theory in Rizzi (1982); two further types are predicted to occur by the theory in Safir (1985). Gilligan tested these predictions on a 100-language sample; he also drew evidence from a smaller survey by van der Auwera (1984). Gilligan observes that fifteen out of the sixteen types are attested.

These two examples testing hypotheses of constructional interactions proposed on the basis of a small number of mostly European languages suggest that in fact there is little or no constraint on the interaction between the constructions. More generally, it implies that the "one language at a time" approach is not a fruitful one for finding syntactic universals. It also makes a typologist very skeptical about the validity of the implicational and biconditional universals implicit in Baker's theory of Pred and parts of speech.

In contrast, the typological approach has discovered an implicational syntactic universal about the presence/absence of a copula in the predication of different parts of speech. The distribution of copulas was of course the initial motivation for Baker's theory of Pred (see above). The implicational universal, a hierarchy, was proposed in Croft (1991:130), and confirmed in Stassen's very large sample (Stassen 1997: 127):[4]

(39) No copula for predicate nominals [predicated object concept words] ⊃ No copula for predicate adjectives [predicated property concept words] ⊃ No copula for predicated verbs [predicated action concept words]

[4] Example (23) is shorthand for a chain of interlocking implicational universals.

This implicational hierarchy was found by examining a broad range of languages simultaneously and has proven to be quite robust. Yet Baker's theory cannot capture this implicational universal, because Baker has dissociated Pred from any actual grammatical form such as an overt copula. (Baker appeals to the Dummy Hypothesis for the distribution of copulas, but as we have seen, that hypothesis is invalid.)

In discussing the problems for the theory of the pro-drop parameter, Gilligan makes some observations that touch on the second serious problem with the generative method for finding syntactic universals. With respect to the Rizzi hypothesis about the possible language types with respect to the pro-drop constructions, Gilligan writes, 'perhaps the Rizzi hypothesis is correct but its effects are obscured in these languages because of some as yet unanalyzed aspect of these languages' (Gilligan 1987: 90). More generally, Gilligan notes, 'as is frequently stated in generative grammar, it is impossible to prove an analysis incorrect; rather, it is only possible to improve upon an existent analysis' (Gilligan 1987: 92).[5] This way of evaluating hypotheses follows directly from the other major feature of the generative method. This is that the generative linguist can arbitrarily choose which constructions are to be invoked to justify a particular analysis or syntactic universal of underlying form such as Pred in a particular language. The constructions chosen do not have to be the same from one language to the next. Moreover, if the construction provides conflicting or anomalous evidence regarding the syntactic universal, one can arbitrarily choose yet another construction, until one finds a construction that gives the result that one is looking for. All these strategies can be found in Baker's analysis of Pred in English, Edo, Chichewa, Mohawk, Hebrew and Quechua discussed above.

This is the second serious problem with the generative method. I describe it as methodological opportunism in Croft (2001): the arbitrary selection of a subset of distributional contexts (constructions) in a language in order to identify a theoretical grammatical category/distinction. As with the "one language at a time" method, methodological opportunism extends at least as far back as the American structuralists (Croft 2001, Chapter 1), so it represents something inherited by generative grammar.[6] Methodological opportunism, as its name implies, is basically an unconstrained method. There is no *a priori* basis for selecting, or limiting the selection of, the constructions whose interaction with the grammatical phenomenon being investigated is deemed relevant.

In describing methodological opportunism to various audiences, generative linguists have asked me whether this is simply a "heuristic" or a "discovery procedure", which is of no serious concern to linguistic theory. If the method of syntactic

---

[5] An anonymous referee points out that 'innumerable analyses have been proved incorrect in generative grammar'. I do not disagree, but as my analysis of Baker's arguments and Gilligan's comments indicate, the proof is not solely on the basis of the sort of syntactic argumentation that is discussed in this chapter (see Section 4).

[6] In his thoughtful history of grammatical theory in the United States, Peter Matthews observes that while the explanatory theory of generative grammar (grammar as a mental representation of tacit knowledge, and UG as an innate capacity) differs dramatically from that of American structuralism, the methodology of grammatical analysis and argumentation remains remarkably similar (Matthews 1993: 149, 153, 211–12).

argumentation—for this is what methodological opportunism is—is discounted in this way, the absence of a critical examination of methodological opportunism in effect shields generative theory from empirical confirmation. And this is what we observe. For example, one cannot empirically disconfirm Baker's syntactic universal of Pred in (14), because, to use Lakatos' (1970: 133) terminology, this universal is surrounded by a 'protective belt' of 'auxiliary hypotheses', namely the universals in (20), (23), (24) and (27), which are the only universals that can be directly empirically confirmed. But the real methodological objection is that the auxiliary hypotheses are protean and proliferate to such a degree that the theory becomes so baroque and so sensitive to empirical counterexamples that it quickly becomes untenable. A generative theorist like Baker, when faced with empirical counterexamples to the auxiliary hypotheses (universals) in (20), (23), (24) and (27), can always adjust these universals thanks to methodological opportunism, leaving the universal of Pred untouched. This is what Gilligan means with regard to not being able to prove an analysis incorrect. In effect, methodological opportunism allows the linguist to posit any underlying construct he or she wishes. If the predictions offered by the theory do not hold, then the linguist may look for other constructions which make the distinction claimed to exist (or unify the phenomena claimed to be the same). Since different constructions can be selected in different languages, there is no constraint imposed by crosslinguistic validity of the distributions of the correlated grammatical constructions. Instead, there is a proliferation of auxiliary hypotheses generated by the grammatical peculiarities of each language examined.

This is not a rigorous approach to the method of argumentation used to support hypothesized syntactic universals. In the final section, I propose a highly rigorous method that is superior to the "one language at a time" method and methodological opportunism.

4 Toward Rigorous Syntactic Argumentation

What would be a truly rigorous method of syntactic argumentation, one that would not be shielded almost completely from empirical disconfirmation? The following principles provide such a method.

(i) Crosslinguistic validation of hypotheses. This is the alternative to the "one language at a time" method. As Stassen's survey of the Dummy Hypothesis and Gilligan's survey of the pro-drop parameter show—the only large-scale crosslinguistic surveys of generative (or structuralist) universals that I am aware of—the universals proposed on the basis of a small number of geographically and genetically similar languages are highly unlikely to be valid. The sample of languages need not be enormous. For example, I identified the implicational hierarchy for overt copulas in (24) using a sample of only twelve languages, and it was later confirmed by Stassen's 410 language sample. What matters is that the sample, even a small one, is representative of a broad range of genetic language families and geographical regions. Sampling methods are discussed in Croft (2003: 19–28) and references cited therein.

The sample must also be potentially extendable (i.e. replicable) and applicable to any language. This requirement is guaranteed by the following principles, which are intended as alternatives to methodological opportunism.

(ii) Use the same construction(s) across languages for argumentation. This principle ensures that you know that you are looking at the same phenomenon in each language you examine. This principle is more difficult to adhere to than is commonly believed, and is elaborated in methodological principles (iv) and (v) below.

(iii) Examine distributional patterns in detail. In examining the constructions you have chosen to compare in each language, don't just use one or two fillers. Principle (iii), which is distributional analysis, is (or was) standard practice in structuralist and at least early generative argumentation. If one restricts oneself to a limited set of distributional patterns for practical reasons, ensure that it is the same set across languages, and do not assume that the results extend to distributions not investigated. For example, in my study of parts of speech (Croft 1991) and Stassen's study of intransitive predication (Stassen 1997), we both restricted ourselves to certain semantic classes of lexical roots; we used the same semantic classes across languages, and did not claim to have identified syntactic universals beyond lexical items of those semantic classes. Finally, take note of any mismatches that are found in the distributional patterns, for example the mismatch between postverbal NP "direct objects" in the active voice vs. subjects in the passive voice in the textbook example in Section 2, or Baker's noting of various mismatches in the predictions his theory makes about the manifestations of Pred in the languages he examined.

(iv) Any additional language-specific interaction invoked to account for an apparent counterexample must also be crosslinguistically validated. This is a crucial difference from methodological opportunism. For example, in the textbook example given above, the grammatical role of "direct object" is being defined in one language, English, by the distribution of subjects in the passive construction in that language: anything that cannot be the subject of the passive is defined not to be a direct object. In order to be methodologically consistent, then we must compare the relevant arguments of the predicates 'eat', 'extend', 'give' etc. in the passive voice in all the languages in the study. Baker's evidence for his syntactic universal for Pred in (14), reviewed in Section 3, invokes many language-specific interactions to account for apparent counterexamples. Each of these therefore implies a syntactic universal (made explicit in Section 3) that must be empirically confirmed or disconfirmed, following (i). Again, as the Stassen and Gilligan studies show, validity without crosslinguistic confirmation is by no means a foregone conclusion.

(v) Use crosslinguistically valid criteria to identify potentially universal patterns. Although I have saved this methodological principle for last, this is perhaps the single most important principle after (i), and is the opposite of methodological opportunism. In methodological opportunism, language-specific criteria—particular constructions and their distribution patterns—are used to identify a theoretical category or distinction which is then assumed to the universal. For example, even many typologists use language-specific criteria to define parts of speech categories such as 'noun', 'verb' and 'adjective', roles such as 'subject', 'direct object', 'subordinate clause', and even categories such as 'word'. The assumption is that such categories

must exist, and all that is necessary is to find constructions in a given language that make the distinction. Except, of course, that other linguists sometimes do not share this assumption: for example, many linguists deny that 'adjectives' exist in many languages, and some linguists deny that 'subjects' exist in some languages. These differences in assumptions are exactly that: they are theoretical assumptions. These assumptions are shielded from empirical disconfirmation by methodological opportunism. A linguist who believes in adjectives can find distributional criteria in any language that distinguishes property concept words ('adjectives') from action concept words ('verbs') or object concept words ('nouns'). Conversely, a linguist who does not believe in adjectives can find distributional criteria that unify property concept words with either action concept words ('adjectives are really a subclass of verbs') or object concept words ('adjectives are really a subclass of nouns'). This is actually no different from holding on to the belief in the syntactic universal for Pred in (14) by using methodological opportunism to find constructions in each language that group together predicate nominals and predicate adjectives as against verbal predication.

The alternative, of course, is to use the same criterion from one language to the next. If one defines 'adjective' according to the type of predication and attributive construction used for property concept words, then one should use exactly those same constructions in every language in order to discover universals about 'adjectives'. Principles (iv) and (v) are clearly related. If one defines 'direct object' according to whether it occurs as the subject of the passive construction, then one should use the passive construction and no other in order to discover universals about 'direct objects'. If one uses secondary predication and the existence of a morphological causative to distinguish 'verbs' from 'non-verbs', as Baker does for certain languages, then one should use the combination of those two constructions and no other in order to discover universals about 'verbs' and 'nonverbs'.

The five methodological principles I have given so far provide a rigorous method of syntactic argumentation that will guarantee the discovery of syntactic universals that are highly likely to be valid. They avoid the pitfalls of the "one language at a time" method and methodological opportunism. The five principles are also the basic principles underlying the typological method, except for (iii), whose roots lie in structuralist and generative theories. It must be said that not all typologists adhere to these methodological principles as rigorously as they should (just as not all generative linguists adhere to principle (iii) as rigorously as they should). In particular, there remains a good deal of methodological opportunism in typology when it comes to the crosslinguistic identification of grammatical categories such as the major parts of speech and the core grammatical relations (see Croft 2001 [Chapter 2], 2005). But there are many excellent examples of rigorous syntactic argumentation in typology that have led to the discovery of empirically robust syntactic universals. They begin with Greenberg's classic paper on word order universals whose fortieth anniversary we are celebrating here (Greenberg 1966a), and also another classic, Keenan and Comrie's papers on the accessibility (grammatical relations) hierarchy (Keenan and Comrie 1977, 1979). They continue with monograph-length studies of high quality such as Stassen's analyses of comparative constructions (Stassen 1985)

and intransitive predication (Stassen 1997); Koptjevskaja-Tamm's analysis of action nominalizations (Koptjevskaja-Tamm 1993), Haspelmath's analysis of indefinite pronouns (Haspelmath 1997), and Cristofaro's analysis of subordination (Cristofaro 2003).

The conclusion to this chapter is that the chief difference between generative and typological syntactic universals can be attributed to the different methods used to identify and justify those universals. In generative theory, the methods used are the "one language at a time" method and what I have called methodological opportunism. I have argued that the "one language at a time" method leads to the positing of syntactic universals that are unlikely to be valid across languages; but that methodological opportunism is so loose that a generative linguist can hold on to those universals by opportunistically selecting constructions in apparently anomalous languages that can be interpreted as supporting the original universal. Indeed, the original universal is also opportunistically established in the language first examined in the "one at a time" method. Anomalous distributional evidence from other constructions is ignored; that is, either the anomalous construction is not examined, or if it is, a further construction is invoked to get around the anomaly. There is an important empirical reason why the opportunistic method is so lax, which will become relevant below. It is because constructions in languages have pervasive mismatches in distributional patterns. Methodological opportunism allows a linguist to evade this fundamental empirical fact of syntactic variability and posit highly general syntactic universals.

Conversely, the strict methodological principles in (i)–(v), if rigorously adhered to as in the studies cited in the paragraph before last, can lead to empirically robust syntactic universals, but they are different types of universals than the syntactic universals posited in generative theory. The chief reason for this is the rigor in the methodology. But if we follow the typological methodology seriously, then certain other consequences follow that I have not made explicit above.

First, it forces the linguist to confront variation and mismatches in distributional patterns in languages. Greenberg faced this empirical fact and came up with the first theoretical construct to identify universals of variation, namely the implicational universal. Greenberg also proposed the implicational hierarchy (Greenberg 1966b), and others following him have posited typological prototypes (e.g. Hopper and Thompson 1980), the semantic map model (Anderson 1982, 1986; Croft et al. 1987; Croft 2003; Haspelmath 2003) and multidimensional scaling (Croft and Poole 2008) to analyze more complex universals of syntactic variation.

Second, it also forces the introduction of function or meaning into the analysis of grammatical form from the beginning of the analysis. Principle (ii) requires the linguist to compare the same constructions across languages. Principle (v) requires the linguist to use crosslinguistically valid criteria. However, the crosslinguistic diversity of grammatical form—another facet of the fundamentally variable nature of syntax—means that "same construction across languages" and "crosslinguistically valid criteria" can only be defined via functional equivalence of grammatical structures, that is, equivalence in meaning and information structure (discourse function). This fact has been recognized from the beginning, by Greenberg (1966a: 74),

and reiterated in typological studies and textbooks concerned with methodology (e.g. Keenan and Comrie 1977: 63; Stassen 1985: 14; Croft 2003: 13–17). Any attempt to define grammatical constructions formally, either in themselves or as criteria for defining grammatical categories such as parts of speech, falls foul of methodological opportunism. Some very basic formal properties can be compared in a crosslinguistically valid fashion, but only in conjunction with the functions they perform:

- *Zero vs. overt coding*: but one needs to identify the function being encoded
- *Behavioral potential*: again, one needs to identify the function being encoded by the grammatical behavior
- *Morphologically bound vs. free*: here one needs a crosslinguistically valid criteria for wordhood; yet there are different, mismatching criteria for phonological wordhood (see e.g. Schiering, Hildebrandt and Bickel 2006; see also Bybee 2001: 30–31)
- *Order of elements*: but those elements must be categorized (e.g. 'Noun' and 'Relative Clause'), and the categories of the elements must ultimately be functionally defined
- *Formal groupings ("constituency")*: but there are different, mismatching criteria here too (see Croft 2001: 185–97; Langacker 1997)
- *Derived structural definitions*: These are structural definitions, but defined in relation a functionally-defined basic structure (Croft 2003: 17–18). For example, one can use a functional definition to identify the basic voice construction of a language, then define other nonbasic voice constructions in the language in terms of their structural contrast to the basic construction, as is done in my typological analysis of passive, inverse and related constructions (Croft 2001, Chapter 8). But a derived structural definition of a construction requires a functionally defined construction as the basis for the structural definition.

Finally, methodological rigor has led a number of typologists, including myself, to construction grammar. This is because construction grammar represents complex syntactic structures as pairings of form (syntactic, morphological, phonological) and function (semantic, information structural, even cultural), which is ideally suited for crosslinguistic comparison and validation of hypotheses. Also, grammatical categories and distinctions are defined distributionally, that is, by the constructions in which they occur (or not). To use a couple of examples from this chapter: if 'direct object' is defined in terms of occurrence in the passive construction, then the universals discovered here are really universals about the passive construction, not some abstract relation of 'direct object'. And if the occurrence of Pred is defined in terms of constructions such as secondary predication, morphological causatives, and so on, then any universals discovered are really universals about those constructions and their interrelationships, not about some abstract category Pred. This has led me to the stronger constructional position that constructions are the basic units of syntactic representation and analysis, which I describe as Radical Construction Grammar (Croft 2001).

In this chapter, I have argued that the differences in the methods of syntactic analysis and argumentation between the generative and typological approaches have led to the different types of syntactic universals proposed in the two approaches. In fact, the differences in the methods have also led to the different sorts of explanatory theories that the two approaches espouse. I have taken a critical view of generative methodology, arguing that it is so loose that one can posit almost any universal of underlying structure without the possibility of empirical disconfirmation. I have argued that one can use a more rigorous method of syntactic analysis and argumentation, and that the more rigorous method is in fact the one used by the best typological research (though not all). Moreover, a truly rigorous method of syntactic analysis leads the linguist to confront the fact of pervasive syntactic diversity and variation, the close relationship between function and form, and ultimately a radical constructional model of grammar, that is, the functional-typological approach to language universals.

References

Anderson, Lloyd B. 1982. The 'perfect' as a universal and as a language-particular category. *Tense-aspect: between semantics and pragmatics*, 227–64. ed. Paul Hopper. Amsterdam: John Benjamins.

Anderson, Lloyd B. 1986. Evidentials, paths of change, and mental maps: typologically regular asymmetries. In *Evidentiality: the linguistic encoding of epistemology*, ed. Wallace Chafe and Johanna Nichols, 273–312. Norwood: Ablex.

Baker, Mark. 2003. *Lexical categories: verbs, nouns and adjectives*. Cambridge: Cambridge University Press.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge: Cambridge Univesity Press.

Comrie, Bernard. 1981/1989. *Language Universals and Linguistic Typology.* Chicago: University of Chicago Press. (2nd edition, 1989.)

Comrie, Bernard. 1984. Language universals and linguistic argumentation: a reply to Coopmans. *Journal of Linguistics* 20.155–63.

Comrie, Bernard. 2003. On explaining language universals. In *The new psychology of language, vol. 2*, ed. Michael Tomasello, 195–209. Mahwah, N.J.: Lawrence Erlbaum Associates.

Coopmans, Peter. 1983. Review of Comrie, *Language universals and linguistic theory. Journal of Linguistics* 19: 455–73.

Coopmans, Peter. 1984. Surface word-order typology and Universal Grammar. *Language* 60: 55–69.

Cristofaro, Sonia. 2003. *Subordination*. Oxford: Oxford University Press.

Croft, William. 1991. *Syntactic categories and grammatical relations: the cognitive organization of information*. Chicago: University of Chicago Press.

Croft, William. 1995. Autonomy and functionalist linguistics. *Language* 71.490–532.

Croft, William. 2001. *Radical Construction Grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2003. *Typology and universals*, 2nd edition. Cambridge: Cambridge University Press.

Croft, William. 2004. Syntactic theories and syntactic methodology: A reply to Seuren. *Journal of Linguistics* 40: 637–54.

Croft, William. 2005. Word classes, parts of speech and syntactic argumentation. *Linguistic Typology* 9: 431–41.

Croft, William. 2007. Modeling grammatical categories: a reply to Aarts. *Studies in Language* 31: 409–30.

Croft, William, Hava Bat-Zeev Shyldkrot, and Suzanne Kemmer. 1987. Diachronic semantic processes in the middle voice. In *Papers from the 7th International Conference on Historical Linguistics*, 179–192. ed. Anna Giacolone Ramat, Onofrio Carruba and Guiliano Bernini. Amsterdam: John Benjamins.

Croft, William and Keith T. Poole. 2008. Inferring universals from grammatical variation: multidimensional scaling for typological analysis. *Theoretical Linguistics* 34:1–37.

Gair, James W. 1970. *Colloquial Sinhalese sentence structure*. The Hague: Mouton.

Gilligan, Gary Martin. 1987. A cross-linguistic approach to the pro-drop parameter. Ph.D. dissertation, University of Southern California.

Greenberg, Joseph H. 1966a. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of grammar*, ed. Joseph H. Greenberg, 2nd edition, 73–113. Cambridge, Mass: MIT Press. (Reprinted in *On language: selected writings of Joseph H. Greenberg*, ed. Keith Denning & Suzanne Kemmer, 40–70. Stanford: Stanford University Press, 1990.)

Greenberg, Joseph H. 1966b. *Language universals, with special reference to feature hierarchies.* (Janua Linguarum, Series Minor 59.) The Hague: Mouton.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin. 2003. The geometry of grammatical meaning: semantic maps and cross-linguistic comparison. In *The new psychology of language*, *vol. 2*, ed. Michael Tomasello, 211–42. Mahwah, N. J.: Lawrence Erlbaum Associates.

Hawkins, John A. 1985. Complementary methods in Universal Grammar: a reply to Coopmans. *Language* 61: 569–87.

Hawkins, John A. 1988. On generative and typological approaches to universal grammar. *Lingua* 74: 85–100.

Hopper, Paul & Sandra A. Thompson. 1980. Transitivity in grammar and discourse. *Language* 56: 251–299.

Jacob, Judith M. 1968. *Introduction to Cambodian*. Oxford: Oxford University Press.

Keenan, Edward L. and Bernard Comrie. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.

Keenan Edward L. and Bernard Comrie, 1979. Data on the noun phrase accessibility hierarchy. *Language* 55: 333–51.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Lakatos, Imre. 1970. The methodology of scientific research programmes. In *Criticism and the Growth of Knowledge*, 91–196. ed. Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.

Langacker, Ronald W. 1997. Constituency, dependency, and conceptual grouping. *Cognitive Linguistics* 8: 1–32.

Laudan, Larry. 1977. *Progress and its problems: towards a theory of scientific growth*. Berkeley and Los Angeles: University of California Press.

Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, Mass.: MIT Press.

Newmeyer, Frederick J. 2005. *Possible and probable languages: a generative perspective on linguistic typology.* Oxford: Oxford University Press.

Rizzi, Luigi. 1982. *Issues in Italian syntax*. Dordrecht: Reidel.

Safir, Kenneth. 1985. *Syntactic chains*. Cambridge: Cambridge University Press.

Schiering, René, Kristine Hildebrandt, and Balthasar Bickel. 2006. Crosslinguistic challenges for the Prosodic Hierarchy: evidence from word domains. Available at http://www.uni-leipzig.de/~bickel/research/papers/ index.html.

Stassen, Leon. 1985. *Comparison and universal grammar*. Oxford: Basil Blackwell.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Oxford University Press.

van der Auwera, Johan. 1984. Subject and non-subject asymmetries in the relativization of embedded NPs. In *Sentential complementation: proceedings of the international conference held at UFSAL, Brussels, June 1983*, 257–69. ed. W. de Geest & Y. Putseys.

The Fundamental Left-Right Asymmetry of Natural Languages*

Guglielmo Cinque

Abstract The article discusses a pervasive left-right asymmetry found in the order of modifiers and functional heads associated with distinct lexical heads. In each case, it is shown that one and the same pattern is involved. The account proposed for such an asymmetry is based on a unique underlying structure for each head and the modifiers and functional heads associated with it, in interaction with independent conditions on phrasal movement.

Keywords Word order · left right asymmetry · phrasal movement

In both the typological and generative literature various left-right asymmetries of natural languages have been discussed; among these, the rightward skewing shown by the location of sentential complements with respect to the verb (Dryer 1980, Hawkins 1988, Section 2.2); the similar rightward skewing of relative clauses with respect to their Head (Hawkins 1988, Section 2.1; Cinque 2005b); the cross-linguistic preference of suffixing over prefixing (Cutler, Hawkins, and Gilligan 1988, Hawkins 1988, Section 2.3, Hawkins and Gilligan 1988); the existence of "unbounded leftward movement" vs. the (virtual) inexistence of "unbounded rightward movement" (Bach 1971, 160f; Bresnan 1972, 42ff; Kayne 1994, 54; Cinque 1996; Hawkins 1998); and the left-right asymmetries in quantifier scope interactions mentioned in Lu (1998, 10fn3).

Here I would like to discuss yet another pervasive left-right asymmetry of natural languages: that found in the ordering of functional modifiers and heads to the left and to the right of a lexical head.

The first glimpse of such an asymmetry is to be found in one of Greenberg's universals, his Universal 20:

"When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite." (Greenberg 1963, 87)

G. Cinque
Ca' Foscari University, Venice

* I wish to thank two anonymous reviewers, and Matthew Dryer, for their comments.

S. Scalise et al. (eds.), *Universals of Language Today*, 165–184,
DOI: 10.1007/978-1-4020-8825-4_9,

The left-right asymmetry implicit in Greenberg's formulation appears more clearly when all the modifiers are on the same side of the noun, as is the case in (1). What we find is that to the left of the noun only one order is possible, while to its right two orders are possible (either the same one or its mirror image).[1]

Order of Demonstratives, numerals, and adjectives (Greenberg 1963, Cinque 1996,2005a)

(1)	a	Dem > Num > A > **N**	(English, Malayalam,. . .)
	b	*A > Num > Dem > **N**	0
	c	**N**> Dem > Num >A	(Abu', Kikuyu,. . .)
	d	**N** > A > Num > Dem	(Gungbe, Thai,. . .)

This is not an isolated property of such modifiers. The same pattern is found with the order of attributive adjectives (2), with the order of adverbs (3), with the order of circumstantial PPs (4), with the order of locative and directional prepositions (5), with the order of Mood, Tense, and Aspect morphemes (6), with the order of auxiliaries (and restructuring verbs) (7), etc.

Consider first the order of attributive adjectives. Restricting ourselves, for convenience, just to adjectives of size, color and nationality among the substantial number of existing classes (see Scott 2002, and references cited there), we find that their order is fixed (if we control for the independent relative clause source of attributive adjectives – see Cinque forthcoming for discussion).

Order of attributive adjectives (not derived from RCs)**:** (Hetzron 1978; Sproat and Shih 1991; Cinque 1994, forthcoming; Plank 2006)

(2)	a	$A_{size} > A_{color} > A_{nationality} >$ **N**	(English, Serbo-Croatian. . .)
	b	*$A_{nationality} > A_{color} > A_{size} >$ **N**	0
	c	**N** > $A_{size} > A_{color} > A_{nationality}$	(Welsh, Irish, Maltese. . .)[2]
	d	**N** > $A_{nationality} > A_{color} > A_{size}$	(Indonesian, Yoruba,. . .)

Similarly, if we take some selection of the many different classes of adverbs that are found within the clause (say, the terminative aspect adverb *no longer*, the completive aspect adverb *completely*, and *always*), we find the same thing:

---

[1] This is in fact a simplification, which however does not affect the thrust of the argument. While the prenominal order is Dem > Num > Adj without exceptions (or virtually so), more possibilities than the two Dem > Num > Adj and Adj > Num > Dem are actually attested postnominally (see (17) below, and Cinque 2005a for an illustration of how they can be derived by different leftward movements).

[2] While the relative order of postnominal adjectives of Size, Color, and Nationality in Welsh is the same as the order of the same adjectives in prenominal position in English (cf. Sproat and Shih 1991,Rouveret 1994, Plank 2006), other adjectives (among which quality, age, the functional adjective *other* and demonstratives) show a (postnominal) order which is the mirror image of the English order (see Willis 2006): N A_{size} A_{color} $A_{nationality}$ A_{age} $A_{quality}$ "other" Dem. If movement of the NP (or phrases containing the NP) rather than head movement is responsible for DP internal orders (Cinque 2005a and forthcoming), this mixture of direct and mirror-image orders of nominal modifiers can be reconciled (pace Willis 2006) with a unique, universal, base structure.

Order of adverbs: (Cinque 1999, 42f, Rakowski and Travis 2000, Pearson 2000)

(3)	a	$Adv_{no\ longer} > Adv_{always} > Adv_{completely} > \mathbf{V}$	(English, Chinese,...)
	b	$*Adv_{completely} > Adv_{always} > Adv_{no\ longer} > \mathbf{V}$	0
	c	$\mathbf{V} > Adv_{no\ longer} > Adv_{always} > Adv_{completely}$	((main clause) German, Italian...)
	d	$\mathbf{V} > Adv_{completely} > Adv_{always} > Adv_{no\ longer}$	(Malagasy, Niuean,...)

This is also what we find with the relative order of circumstantial PPs. If we limit ourselves to Time, Place and Manner PPs, whose order has been investigated from a cross-linguistic perspective by Boisson (1981), and Lu (n.d.) (also see Cinque 2002, Hinterhölzl 2002, Schweikert 2005), we find the same pattern:[3]

Order of circumstantial PPs

(4)	a	Time > Place > Manner **V**	(Basque, Nambikuara,.. – Lu n.d., Kroeker 2001, 3)
	b	*Manner > Place > Time > **V**	0
	c	**V** > Time > Place > Manner	(V/2 clause German)
	d	**V** > Manner > Place > Time	(Vietnamese, Yoruba – Lu n.d.)

A similar pattern is apparently found (in those languages in which they overtly combine) with the order of locative ('at') and directional ('to', 'from') prepositions:[4]

Order of directional and locative prepositions

(5)	a	$P_{Dir}\ P_{Loc}$ NP	(Romanian: *Ion vine de la şcoală* '(lit.) Ion comes from at school (from school)' – Zegrean 2007, 79)
	b	$*P_{Loc}\ P_{Dir}$ NP	0
	c	NP $P_{Dir}\ P_{Loc}$	(Iatmul (Papuan): *gay-at-ba* '(lit.) house-to-at (to the house)' – Staalsen 1965, 21)
	d	NP $P_{Loc}\ P_{Dir}$	(Jero (Tibeto-Burman): *thalu=na=k* 'where=LOC=SOURCE (from where)' – Opgenort 2005, 92)

This is also what we find with the order of (speech act) Mood, Tense, and Aspect with respect to the V (see Bybee 1985, Foley and Van Valin 1984, Cinque 1999, 2007, and the text below):

Order of (speech act) Mood, Tense, and Aspect morphemes

(6)	a	Mood Tense Aspect **V**	(Nama, Yoruba,...)
	b	*Aspect Tense Mood **V**	0
	c	**V** Mood Tense Aspect	(Comox,..)
	d	**V** Aspect Tense Mood	(Korean, Malayalam,...)

[3] On the interference of focus on the canonical order of circumstantial PPs and possible diagnostics for the canonical order, see Cinque (2002), Schweikert (2005).

[4] The other two possible orders of the three elements $P_{Dir}\ P_{Loc}$ NP are also attested: P_{Dir} NP P_{Loc} in Taba (Austronesian - Bowden n.d. *ap-po bbuk li* '(lit.) to-down book at' (onto the book)), and P_{Loc} NP P_{Dir} in Zina Kotoko (Chadic-Tourneux 2003: 294 *'à jì kàskú kí* 'LOC inside market toward' (toward the market)).

If one considers the relative order of auxiliary and restructuring (or clause union) verbs (Cinque 2006) with respect to each other and to the lexical verb, one finds a similar pattern. See Koopman and Szabolcsi (2000), Nilsen and Vinokurova (2000), Wurmbrand (2004), Barbiers (2005), and Svenonius (2006):

Order of auxiliary (restructuring) verbs

(7) a Aux_1 Aux_2 Aux_3 **V** (Italian, English,...)
b *Aux_3 Aux_2 Aux_1 **V** 0
c **V** Aux_1 Aux_2 Aux_3 (Hungarian, West Flemish,...)
d **V** Aux_3 Aux_2 Aux_1 (Hungarian, German,...)

The same pattern is also found within a single language, with respect to the ordering of certain elements. To take one example, Terzi (1999) notes that in front of the verb in Modern Greek only the order in which the dative clitic precedes the accusative clitic is admitted, while after the V either order of the two clitics is possible (see (8)):

Order of (dative and accusative) clitics in Modern Greek (Terzi 1999, 86)

(8) a mou to edoses
me_{dat} it_{Acc} gave.2sg
'you gave it to me'
b *to mou edoses
it_{Acc} me_{dat} gave.2sg
c Dos' mou to
give me_{dat} it_{Acc}
'give it to me!'
d Dos' to mou
give it_{Acc} me_{dat}
'give it to me!'

All of the cases seen above instantiate exactly the same pattern:

(9) a AB(C)X°
b *(C)BA X°
c X° AB(C)
d X° (C)BA

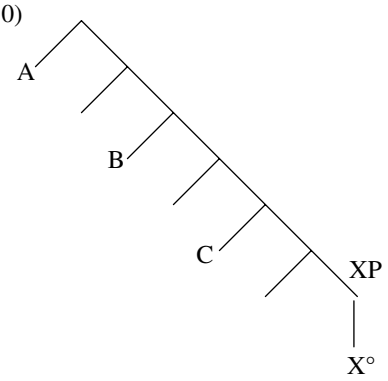
Clearly, this cannot be an accident. It is equally clear that these orders are not independent of one another. One feels in fact that they are the *same* order at a more abstract level, for they are either literally the same, modulo their pre- or post-head location ((9)a and c), or the mirror image of each other on the two sides of the head ((9)a and d). It would thus seem desirable to express this more abstract identity by deriving them from a unique structure.

Sometimes it is assumed that this more abstract identity is expressed by a principle which determines the relative distance of each class of elements from the head, thus accounting for what are possibly the two most common orders of each of the

above cases, (9)a (ABC X°) and (9)d (X° CBA), and for the non existence of the order (9)b (CBA X°). But, if one takes this line, one can only state the principle as a tendency given that the fourth order, (9)c (X° ABC), even if it is generally rarer, plainly violates it.

The principle (whatever it ultimately follows from) can however be stated as an absolute principle, rather than just a tendency, if we are willing to abandon the symmetrical view underlying the above account (as in fact Kayne's 1994 antisymmetry principle would have us do), and to adopt a more abstract, asymmetrical, view, whereby there is only one order/structure available for all languages (10), and whatever word order difference there is among them is a function of independently motivated types of movement of the lexical core XP.

(10)



We know that certain phrases in certain languages can, or must, appear displaced; for example (single) interrogative *wh*-phrases in English must be displaced to sentence initial position (as in (11), below). And we know that languages vary with respect to whether they displace them or not. In some languages (e.g., Indonesian - see (12)) *wh*-phrases remain *in situ*. We also know that depending on certain conditions movement can affect just the phrase bearing the feature triggering the movement - here the *wh*-feature – (as in (11)), or a larger phrase containing the phrase bearing the relevant feature (as in (13)); what Ross (1967) called Pied Piping:

(11) [*Who*] did you see ?

(12) Siti mau apa? (Cole, Hermon and Tjung 2005, 553)
Siti want what
'What does Siti want?'

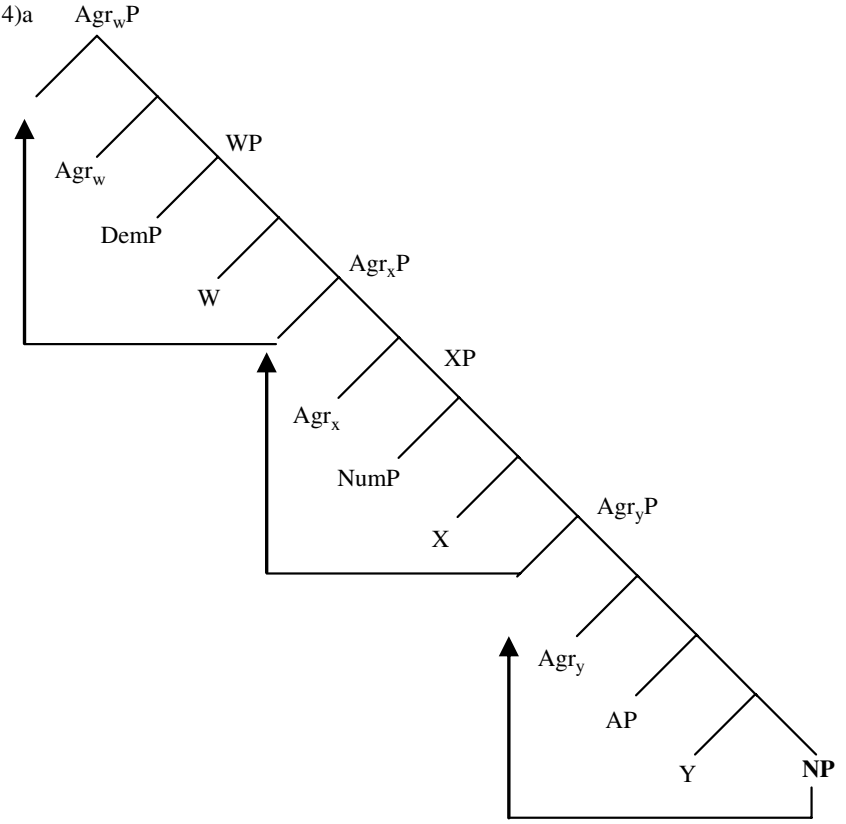
(13) [[*Wh*ose] pictures] did you see ?

In Cinque (1996, 2003, 2005a) I suggested that precisely these two independent parameters (whether the relevant phrase remains in situ or moves; and, if it moves, whether it moves by itself, or by pied piping each time the immediately dominating phrase) can account for the three attested orders of Dem Num A N ((1)a,c,d) and for the principled absence of the fourth ((1)b).

The phrase bearing the relevant feature triggering the movement (a nominal feature) is in this case NP.

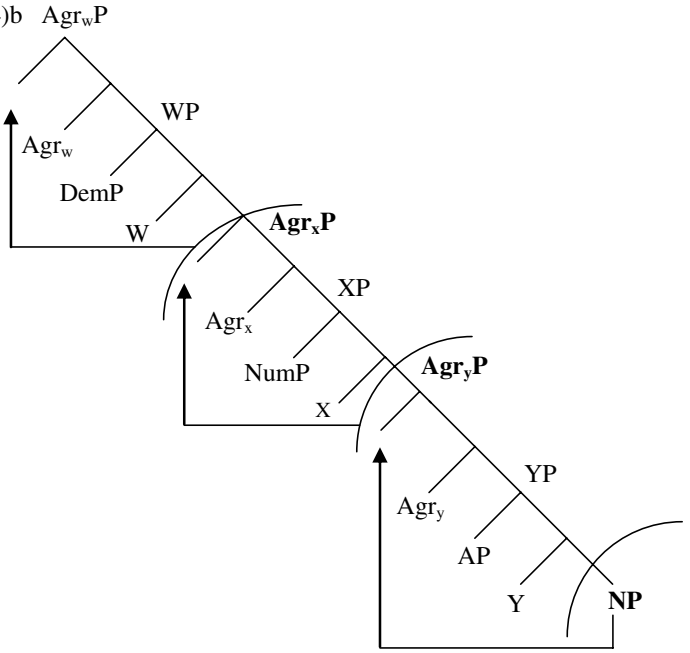
If NP does not move, we get (1)a. If NP moves by itself (all the way up), as shown in (14a), we get (1)c. If it moves (all the way up) each time pied piping the immediately dominating phrase, as in (14b), we get (1)d. (1)b cannot be derived because the NP has not moved and the base structure has the modifiers in the wrong order. Crucially AP, NumP, or DemP cannot move by themselves just as phrases not bearing the *wh*-feature cannot move by themselves to the sentence initial +*wh*-position.[5]

(14)a



[5] In certain languages, (at most) one of these elements, if it bears a focus feature, can apparently move to an initial focus position– see fn.21 below for relevant references.

(14)b



Note that if the principle governing the degree of proximity of each modifier to the head is stated on the "base level" (10), before movement takes place which disrupts the original order of elements, it can be stated as an absolute principle forcing AP to be merged closer to the head than NumP, and NumP closer to the head than DemP.

This logic extends to the other instances of the same pattern seen above.

This is however a simplification. The orders that it accounts for are the orders in (1)a,c,d, repeated here as (15)a–c, and, taking partial movement into account (i.e., when the NP does not move all the way up), the orders in (16)a–c:

(15) a Dem Num A N
b N Dem Num A
c N A Num Dem

(16) a Dem Num N A
b Dem N Num A
c Dem N A Num

But, of the 24 mathematically possible orders of the four elements Dem, Num, A and N, more than the six indicated in (15) and (16) are attested, as is apparent from the table in (17), from Cinque (2005a), which documents 14 orders as attested (although in the same article I suggested that one ((17)r) may be spurious, with the position of A really being the position of reduced relative clauses).[6]

[6] The references in the footnotes that follow are those given in Cinque (2005a), with some additions.

(17)

a.	√	Dem	Num	A	**N**	(very many languages)[7]
b.	√	Dem	Num	N	A	(many languages)[8]
c.	√	Dem	**N**	Num	A	(very few languages)[9]
d.	√	**N**	Dem	Num	A	(few languages)[10]
e.	*	Num	Dem	A	**N**	(Ø – Greenberg 1963; Hawkins 1983)
f.	*	Num	Dem	**N**	A	(Ø – Greenberg 1963; Hawkins 1983)
g.	*	Num	**N**	Dem	A	(Ø – cf. Lu 1998, 183)
h.	*	**N**	Num	Dem	A	(Ø – cf. Greenberg 1963; Lu 1998, 162)
i.	*	A	Dem	Num	**N**	(Ø – Greenberg 1963; Hawkins 1983)
l.	*	A	Dem	**N**	Num	(Ø – Greenberg 1963; Hawkins 1983)
m.	√	A	**N**	Dem	Num	(very few languages)[11]
n.	√	**N**	A	Dem	Num	(few languages)[12]

[7] Rijkhoff (1998, 357) states that the "order [Dem Num A N] is by far the most common both inside and (to a lesser extent) outside Europe", listing on p. 342f many languages of the Afro-Asiatic, Altaic, Caucasian, Indo-European, and Uralic families. More languages with this order are listed in Hawkins (1983, 119), Rijkhoff (1990, 32; 2002, 112, 270, fn.10, 310, 328, 330f), and Croft and Deligianni (2001, 7). It is also found in Amerindian (e.g., Comox – Harris 1977, 129) and Australian (e.g. Tiwi – Osborne 1974, 73) languages.

[8] According to Rijkhoff (1998, 357) "[t]he order [Dem Num N A] is [..] rather frequent in Europe". Outside Europe it is documented, among other languages, in Yao (Jones 1970), Khasi (Nagaraja 1985, 14ff), Madak (Lee 1994, Section 1.1), Burushaski, Guaraní (Rijkhoff 2002, 328), Abkhaz, Farsi, Kiowa, Mam (Croft and Deligianni 2001), Kristang, Kriyol, Tok Pisin and Cape Verdian, Mauritian, and Seychelles Creoles (Haddican 2002).

[9] This order is documented in Sampur and Camus (Heine 1981) (but see Rijkhoff 2002, 274f), in Maasai (Koopman 2003), and in Wappo (Thompson, Park, Li, 2006, 8). According to Croft and Deligianni (2001, 7), it is also a possible alternative order (of the Dem N A Num order) in Hualapai and Lahu.

[10] Greenberg (1963, 87) states that the N Dem Num A is "[a] less popular alternative" to N A Num Dem, citing Kikuyu as one example. Other languages displaying this order are: Elmolo (Heine 1980), Turkana, Rendille (Heine 1981) Noni (or Nooni – Hyman 1981, 31; Lux and Lux 1996, 10), Nkore-Kiga (Lu 1998, 162fn59, 165), Nomaándɛ́ (Wilkendorf n.d., 11), Abu' (Lynch 1998, 171), Arbore (Hayward 1984, 212), Bai and Moro (Dryer 2007, 20 and 43), and the Kuliak (Nilo-Saharan) languages Ik and So (Serzisko 1989, 391). This is also the order given by Lawton (1993, 150) for Kiriwina (Kiliwila).

[11] It is found in Koiari (which also has the order N A Dem Num with most adjectives – Dutton 1996, 60ff), and in Bai (Wiersma 2003, 669). According to Dryer (2000, 20), Bai also has N Dem Num A as an alternative order. [A N]-def Num is also an alternative order of the unmarked Dem Num A N order of Icelandic (Sigurðsson 1993, 194; Vangsnes et al. 2004). The possibility of this order in Koiari, and Bai (and of the order A N Num Dem in Gude and Sango – see below) indicates that the last sentence of Hawkins' (1983, 119–120) revision of Greenberg's Universal 20 ("In no case does the adjective precede the head when the demonstrative or numeral follow.") may be too strong. Greenberg's (1963) Universal 18 was less categorical ("When the descriptive adjective precedes the noun, the demonstrative, and the numeral, with overwhelmingly more than chance frequency, do likewise"). This was because of the existence, noted by Greenberg, of "a small number of instances (e.g., Efik) in which the demonstrative follows while the adjective precedes" (p. 86). Cf. also Dryer (2000, 34).

[12] This order is found in Lalo (Björverud 1998, 116ff), Lisu (Bradley 2003, 228f), Akha (Hansson 2003, 241), Aghem (Hyman et al. 1979, 27), Maranunggu (Tryon 1974, 154), Kenyang (Ramirez 1998, 28), Port Sandwich (Crowley 2002, 653), Koiari (Dutton 1996, 60ff), which also has the order A N Dem Num with certain adjectives, Lingala (Haddican 2002), Hocank, which also has the alternative order N A Num Dem (Helmbrecht 2004, 13). Croft and Deligianni (2001) also assign to this order Babungo and , more tentatively, Woleaian.

o.	*	Dem	A	Num	**N**	(Ø – Greenberg 1963; Hawkins 1983)[13]
p.	√	Dem	A	**N**	Num	(very few languages)[14]
q.	√	Dem	**N**	A	Num	(many languages)[15]
r.	√	**N**	Dem	A	Num	(possibly spurious)
s.	*	Num	A	Dem	**N**	(Ø – Greenberg 1963; Hawkins 1983)
t.	√	Num	A	**N**	Dem	(very few languages)[16]
u.	√	Num	**N**	A	Dem	(few languages)[17]
v	√	**N**	Num	A	Dem	(few languages)[18]
w.	*	A	Num	Dem	**N**	(Ø – Greenberg 1963; Hawkins 1983)
x.	*	A	Num	**N**	Dem	(Ø – Greenberg 1963; Hawkins 1983)
y.	√	A	**N**	Num	Dem	(very few languages)[19]
z.	√	**N**	A	Num	Dem	(very many languages)[20]

[13] A potential counterexample, pointed out to me by Matthew Dryer (p.c.), is provided by Dhivehi (Maldivian), for which Cain (2000, 78), and Cain and Gair (2000, 33) give Dem A Num N as the canonical order. Whether this exception is a real counterexample or can be explained away by assuming that Dhivehi lacks direct modification (i.e., non relative clause derived) adjectives entirely, and exploits the possibility of introducing them as the predicate of a (prenominal) reduced relative clause (like possibly in (17)r, as noted) will be left open here.

[14] According to Hawkins (1983, 119), Lu (1998, 165), and Rijkhoff (1998, 358; 2002, 331), this order is not attested. However, Kölver (1978, 285) documents it in Newari (also see Dryer's (2000, 39) example (79)), Lapolla (2003, 676) in Dulong, Mazaudon (2003, 297) in Tamang, Gair and Paolillo (1997,29f) in Sinhala, and Valenzuela (2002,28f) in Shipibo-Konibo. Bhattacharya (1998) and Croft and Deligianni (2001) give it as an alternative order for the Dem Num A N order in, respectively, Bangla (where it leads to a specific interpretation of the DP) and Syrian Arabic.

[15] Among the languages that instantiate this order are Kabardian and Warao (Hawkins 1983, 119; Colarusso 1992, 63), Burmese, Lolo, Maru, Răwang (Jones 1970), Manange (Genetti and Hildebrandt 2004, 75), Ladakhi (Koshal 1979, 108), Epena Pedee (Harms 1994, Chapter 4), Miya (Schuh 1998, 277), Gambian Mandinka (Rijkhoff 1998, 356), Cuna (Quesada 1999, 232), Kaki Ae (Clifton 1995, 46), Pech (Holt 1999, 62ff), Tunen (Mous 1997, 124). It is an alternative order of N A Num Dem in Kunama (Bender 1996, 41), and of Dem N Num A in Hualapai and Lahu (Croft and Deligianni 2001, 7).

[16] According to Hawkins (1983, 119) and Lu (1998, 165) this order is not attested. However, Rijkhoff (2002, 328) reports Berbice Dutch Creole as instantiating it. Haddican (2002) documents the same order for the Creole language Bislama. Lynch (2002, 769f,781,809) gives it as the order of Xârâcùù, Iaai, and Puluwatese. To judge from Siewierska and Uhlířová (1997, 132f), Polish and Russian also have this order as an alternative order to Dem Num A N.

[17] This order appears documented in a number of Mon-Khmer languages (Dryer 2001), in Basque (Rijkhoff 2002, 328), Celtic, Easter Island, Hebrew, Indonesian, Hmong, Jacaltec, Miao (cf. Hawkins 1983, 119, Lu 1998, 162; Harriehausen 1990, 144), in Nung (Saul and Freiberger Wilson 1980, 14), in Vietnamese (Nguyen 2004) in Wolof (Sy 2003), in Sisiqa (Ross 2002a, 459f); and in a number of Creoles (Haddican 2002). It is also displayed by the Australian language Watjarri (Douglas 1981, 241).

[18] According to Lu (1998, 162) this order is not attested. However, Heine (1981), as noted, documents it in three languages: Gabra, Logoli and Luo (on Luo, also see Chiao 1998). Noonan (1992, 154) documents it in Lango. Ross (2002a, 132) and Tryon (2002, 576) give it as the order of Kele, and Buma, respectively. Croft and Deligianni (2001) give it as an alternative order in Manam.

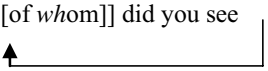
[19] According to Hawkins (1983, 119) and Lu (1998, 165), this order is not attested. However, Thornell (1997, 71) and Haddican (2002) give it as the order of Sango and Rijkhoff (1998, 356, 358; 2002, 332, fn.19) mentions (dubitatively) the possible existence of two other languages with this order: Gude and Zande.

[20] Cambodian, Javanese, Karen, Khmu, Palaung, Shan, Thai (Rijkhoff 1990, 32), Enga (Lynch 1998, 171), Dagaare (Bodomo 1993), Ewe (Essegbey 1993), Gungbe (Aboh et al. 2004),

All of the attested orders, and none of the unattested ones, can be derived, it seems, by slightly refining our earlier assumptions.

Note that in addition to the Pied Piping of the [[wh*ose*] *pictures*] type, which drags along constituents to the right of the phrase triggering movement, there is also a Pied piping of the [*pictures* [*of* wh*om*]] type, which drags along constituents to the left of the phrase triggering movement:

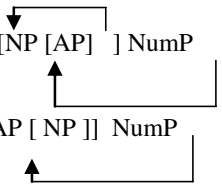
(18) [pictures [of *wh*om]] did you see ?



This means that in addition to movements like the one in (19)a, giving the order N A Num, one can also expect to find movements like the one in (19)b, giving the order A N Num:

(19)a ..[NP [AP]] NumP

b [AP [NP]] NumP



As I suggested in (2005a), all of the attested orders (and none of the unattested ones) can be derived if we revise our earlier assumptions in the way indicated in (20):

(20) a. Base order: [. . . [$_{WP}$DemP . . . [$_{XP}$ NumP . . . [$_{YP}$AP [$_{NP}$N]]]]]
b. Parameters of movement:
i) No movement (unmarked), or
ii) NP movement plus Pied-piping of the *whose pictures*-type (unmarked), or
iii) NP movement without Pied-piping (marked), or
iv) NP movement plus Pied-piping of the *pictures of whom*-type (more marked still)
v) *total* (unmarked) vs. *partial* (marked) movement of the NP with or without Pied-piping (in other words, the NP raises all the way up, or just partially, around its modifiers).
vi) Neither head movement nor movement of a phrase not containing the NP are possible (except perhaps for a single focus-related movement to a DP initial position).[21]

Labu and Ponapean (Lynch 1998, 121), Mao Naga (Giridhar 1994, 452) Selepet, Yoruba (Hawkins 1983, 119), West Greenlandic (which also has N A Dem Num as an alternative order) (Rijkhoff 2002, 326); Amele, Igbo, Kusaeian, Manam (Croft and Deligianni 2001), Fa d'Ambu, Nubi (Haddican 2002), Kugu Nganhcara (Smith and Johnson 2000, 388), Cabécar (Quesada 1999, 232), Kunama (Bender 1996, 41), Māori (Pearce 2002).

[21] On the possible, marked, preposing of APs to DP initial position (often for focusing purposes), see Corbett (1979), Giusti (1996), and Rijkhoff (1998, 352f; 2002, 267, 272). One additional parameter is the *obligatory* vs. *optional* application of movement. For example, the alternative orders

The "marked", "unmarked", "more marked", etc., values attached to each parameter of movement (some of which appear to be independently motivated – see Cinque 2005a) were meant to account, at least in part, for the different numbers of languages that appear to instantiate the different orders (although no precise statistics were carried out).

I review here the derivation of some of the orders in (17) (for a systematic review of all of the orders see Cinque 2005a).

a. (Dem Num A **N**) is derived if nothing moves (no marked option: very many languages).
d. (**N** Dem Num A) is derived if NP moves three notches, around A, Num, and Dem (i.e. all the way up) without Pied-piping (one marked option: few languages).
e. (Num Dem A **N**) cannot be derived. NP has not moved, and the modifiers to its left are in the wrong order of Merge.
m. (A **N** Dem Num) has a well-formed, though marked, derivation with raising of NP plus Pied-piping of the *pictures of whom*-type of the lowest modifier (A) around Num, followed by raising (of [A N]) without Pied-piping around Dem (two marked options: very few languages)
n. (**N** A Dem Num) has a derivation with NP raising past A, followed by Pied-piping of the *whose pictures*-type past Num, followed by raising (of [N A]) without Pied-piping (marked) past Dem (one marked option: few languages).
p. (Dem A **N** Num) has a derivation with partial (marked) raising of NP plus Pied-piping of the *pictures of whom*-type of [A N] (marked) around Num (two marked options: very few languages)
t. (Num A **N** Dem) has a derivation with partial (marked) raising of NP plus Pied-piping of the *pictures of whom*-type of A and Num ([Num A N]) (marked) around Dem (two marked options: very few languages).

The question that arises is whether exactly the same fine-grained variation that we find with the order of Dem Num A and N is also found with the order of the other elements reviewed in (3)–(7). I think it is.

In Cinque (2007), I documented it for the relative orders of (speech act) Mood, Tense, Aspect and V. The order of these elements is often taken to be governed by a principle that determines the degree of proximity to the V of Mood, Tense, and Aspect morphemes (Aspect being closer to V than Tense, which in turn is closer to V than speech act Mood – see Gerdts' 1982,193fn4 "Satellite Principle", Bybee's 1985 "Principle of Relevance", Foley and Van Valin 1984 "Principle of Scope Assignment", and Baker's 1985 "Mirror Principle").

Q Dem Num **N** A, Q Dem **N** A Num, Q **N** A Num Dem, **N** A Num Dem Q of Standard Arabic (cf. Fassi Fehri 1999) point to the obligatory character of movement of the NP around the adjectives followed by optional movements (plus Pied-piping of the *whose picture*-type) around numerals, demonstratives and universal quantifiers.

These principles account for the two prevailing orders of such elements ((21)a–b), but, as shown in table (22), the actual orders attested are thirteen, five of which (c.,d.,m.,n.,v.) do not conform to the proposed principles.[22]

(21) a. Mood Tense Aspect V
b. V Aspect Tense Mood

(22)

a.	√	Mood	Tns	Asp	**V** [23]	
b.	√	Mood	Tns	**V**	Asp [24]	
c.	√	Mood	**V**	Tns	Asp [25]	
d.	√	**V**	Mood	Tns	Asp [26]	
e.	*	Tns	Mood	Asp	**V**	(Ø)
f.	*	Tns	Mood	**V**	Asp	(Ø)
g.	*	Tns	**V**	Mood	Asp	(Ø)
h.	*	**V**	Tns	Mood	Asp	(Ø)
i.	*	Asp	Mood	Tns	**V**	(Ø)[27]
l.	*	Asp	Mood	**V**	Tns	(Ø)
m.	√	Asp	**V**	Mood	Tns [28]	
n.	√	**V**	Asp	Mood	Tns [29]	

[22] Sources documenting the attested orders are given in the footnotes that follow. See Cinque (2007) for examples illustrating the various orders, and discussion on some apparent exceptions.

[23] This order is attested in Khoisan (e.g., Nama: http://instruct1.cit.cornell.edu/courses/ling700/nama.htm, and /Xam http://instruct1.cit.cornell.edu/courses/ling700/xam.htm); in a number of Oceanic (Austronesian) languages (‘Ala‘ala - Ross 2002c, 353 and 359; Nabukelevu – Pawley and Sayaba 1982, 68 and 85; Samoan - Cinque 1999, 160); in Yoruba (Niger-Congo – Ọládiípò Ajíbóyè, p.c.); and in some South American Indian languages (Apinajé (Macro-Jê) – Cunha de Oliveira 2003, 255f, 265), and Canela–Crahô (Cariban – cf. Cinque 1999, 162 and references cited there).

[24] In addition to Nama (which also instantiates the order in (22)a), and N|uu (Khoisan - Collins 2004, 188), other languages instantiating this order are Easter Island (Austronesian - Chapin 1978, 153, 168), Hmong Njua (Sino- Tibetan - Harriehausen 1990, 57, 226); and Nabukelevu (with postverbal progressive aspect markers – Pawley and Sayaba 1982, 53ff).

[25] This order is found in, among other languages, Kharia (Munda-Biligiri 1965, 59, 98), Ngarinjin (Kimberley, North Western Australia – Coate and Coate 1970, 43, 75), and Tümpisa Shoshone (Uto-Aztecan - Dayley 1989, 325, 348).

[26] This order appears instantiated in Comox (Central Coast Salish - Harris 1977, 139), and, to judge from Aikhenvald (2006, 179, 190) (at least for some combinations of Mood, Tense and Aspect) in Tariana (North Arawak).

[27] St’át’imcets (Matthewson 2003, 69) apparently shows the order imperfect > interrogative > past > V, but the interrogative particle is a second position particle, with the imperfect particle possibly moved to first position from a lower one (see the discussion in Cinque 2007).

[28] This order appears to be instantiated in Xârâcùù (Moyse-Faurie 1995, 117, 157), and Tinrin (Osumi 1995, 188, 204), two Melanesian (Austronesian) languages of New Caledonia, and in Sooke (Coast Salish - Efrat 1969, 43, 189).

[29] This order is instantiated in Kanoê (a language isolate of Brasil) with Past tense (Bacelar 2004, 222, 226), in Lummi (Coast Salish - Steele 1981, 60; and Jelinek and Demers 1997, 310f), and in Lotha (Tibeto-Burman – Acharya 1983, 158).

o.	*	Mood	Asp	Tns	**V**	(Ø)
p.	√	Mood	Asp	**V**	Tns [30]	
q.	√	Mood	**V**	Asp	Tns [31]	
r.	*	**V**	Mood	Asp	Tns	(Ø)
s.	*	Tns	Asp	Mood	**V**	(Ø)
t.	√	Tns	Asp	**V**	Mood [32]	
u.	√	Tns	**V**	Asp	Mood [33]	
v	√	**V**	Tns	Asp	Mood [34]	
w.	*	Asp	Tns	Mood	**V**	(Ø)
x.	*	Asp	Tns	**V**	Mood	(Ø)
y.	√	Asp	**V**	Tns	Mood [35]	
z.	√	**V**	Asp	Tns	Mood[36]	

The same parameters (with VP in place of NP) that we saw in (20) appear to provide an account of the attested and unattested orders of Mood, Tense and Aspect with respect to the verb.

[30] This order is documented in Gunwinggu, a North Australian language of Arnhem Land (Oates 1964, 49, 53, 82), and in Nevome (Uto-Aztecan – Shaul 1986, 25, 85). It also appears to be instantiated in Slave (Athapaskan – Rice 1989, 420, 588, 1003).

[31] This order is documented in, among other languages, Santali (Munda - Gosh 1994, 106, 152), Northern Pomo (Hokan - O'Connor 1992, 47, 269), Iatmul (Papuan – Staalsen 1972, 49, 50, 57), and in the Australian languages Gidabal (Geytenbeek and Geytenbeek 1971, 45) and Pitjantjatjara (Glass and Hackett 1970, 32 and 74).

[32] This order appears to be instantiated in a number of Austronesian languages, among which Loniu (Hamel 1994, 149) and Tigak (Beaumont 1979, 35 and 78ff). It is also displayed by Kom (Benue-Congo – Chia 1976), Blackfoot (Algonquian - Frantz et al. 1991), Sm'algyax (Penutian – Mulder 1994, 80, 178),and Cogtse Gyarong (Tibeto-Burman – Nagano 2003, 476f).

[33] This order appears to be instantiated in a number of Oceanic (Austronesian) languages, among which Kairiru (Ross 2002b, 211, 214), Kaulong (Ross 2002d, 400, 409), and Urak Lawoi' (Hogan 1999, 38, 40).

[34] Fernandez (1967, 30 and 44) explicitly claims that this is the order of tense, aspect, and interrogative mood suffixes in Remo (Munda-Khmer). The same order is apparently attested in the Niger-Congo languages Mundang (Adamawa - Elders 2000, 387, 389) and Noon (West Atlantic – Soukka 2000, 181, 200), and in Creek (Muskogean – Martin 2000, 388). It is also documented in a number of Tibeto-Burman languages (e.g., Limbu -Van Driem (1987, 90); and Apatani - Abraham 1985, 95, 103).

[35] This order is instantiated in a number of (non-Austronesian) Papuan languages of New Guinea: Amanab (Minch 1991, 10, 17ff, 60), Namia (Feldpausch and Feldpausch 1992, 55), Nend (Harris 1990, 139 and 154), Yagaria (Renck 1975, 101); in the Austronesian languages Urak Lawoi' (Hogan 1999, 7f and 19), in Diegueño (Hokan - Langdon 1970, 147 and 186), in Slave (Athapaskan - Rice 1989, 1114, 1131). This order is also found with free morphemes in Tondi Songway Kiini (Nilo-Saharan - Heath 2005, 175, 182), and Mina (Chadic - Frajzyngier and Johnston 2005, 183, 200).

[36] This is by far the most frequent order. It is typical of Altaic, Caucasian, Dravidian, Eskimo-Aleut, Manchu-Tungusic, Tibeto-Burman, and Papuan languages, and it is also found in many Amerindian, and Indo-European, languages.

Barbiers (2005) shows that much the same holds for the orders of two auxiliary/modal verbs and the lexical verb attested in the dialects of Dutch.

What remains to be seen is whether the rest of the patterns of (3)–(7) also show the same variation displayed by Dem Num A N and Mood Tense Aspect V. If they do, there will not only be evidence for the existence of the left-right asymmetry discussed here, but also some plausibility to the idea that such asymmetry should be accounted for in terms of a unique hierarchical structure shared by all languages, with extant differences stemming from the limited (and independently motivated) ways phrases can move. This is because such an account can discriminate precisely between the actually attested orders and the unattested ones.

A more general implication of this analysis, if correct, is that the lexical head is the lowest head of the projection (the one starting the syntactic computation), and that constituents found to the right of the lexical head are not base-generated there, but come to be there as a consequence of the head moving leftward past them, merged in pre-head position. Only if we assume that can we provide a unique structure underlying all attested word order variations in terms of independently motivated types of movement.

References

Aboh, Enoch O. 2004. *The Morphosyntax of Complement-Head Sequences. Clause Structure and Word Order Patterns in Kwa.* New York: Oxford University Press.

Abraham, P.T. 1985. *Apatani Grammar*. Manasagangotri: Central Institute of Indian Languages.

Acharya, K.P. 1983. *Lotha Grammar*. Manasagangotri: Central Institute of Indian Languages.

Aikhenvald, A.Y. 2006. Serial Verb Constructions in Tariana. In *Serial Verb Constructions. A Cross-linguistic Typology*, eds., A.Y. Aikhenvald and R.M.W. Dixon. 178–20. Oxford: Oxford University Press.

Bacelar, L.N. 2004. *Gramática da língua Kanoê*. Nijmegen: Katholieke Universiteit Nijmegen.

Bach, E. 1971. Questions. *Linguistic Inquiry* 2: 153–166.

Baker, M. 1985. The Mirror Principle and morphosyntactic explanation. *Linguistic Inquiry* 16: 373–415.

Barbiers, S. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In *Syntax and Variation*, eds. L. Cornips and K.P. Corrigan, 233–264. Amsterdam: Benjamins.

Beaumont, C.H. 1979. *The Tigak Language of New Ireland*. Canberra: Australian National University (Pacific Linguistics, Series B-58).

Bender, L.M. 1996. *Kunama*. München: Lincom Europa.

Bhattacharya, T. 1998. Specificity in the Bangla DP. In *Yearbook of South Asian Languages and Linguistics,* vol. 2., ed. R. Singh, 71–99. New Delhi/London: Sage Publications.

Biligiri, H.S. 1965. *Kharia.Phonology, Grammar and Vocabulary*. Pune: Deccan College.

Björverud, S. 1998. *A Grammar of Lalo.* Lund: Department of East Asian Languages (University of Lund).

Bodomo, A.B. 1993. Dagaare Syntax: A Two Level X-Bar Account. *University of Trondheim Working Papers in Linguistics* 19: 20–51.

Boisson, C. 1981. Hiérarchie universelle des spécifications de temps, de lieu, et de manière. *Confluents* 7: 69–124.

Bowden, J. n.d. Taba as a Split-O Language: applicatives in a split-S system. *Research School of Pacific and Asian Studies Working Papers in Linguistics* (available at: http://rspas.anu.edu.au/linguistics/projects/WP/Bowden2.html).

Bradley, D. 2003. Lisu. In *The Sino-Tibetan Languages*, eds. G. Thurgood and R. LaPolla. 222–235. London: Routledge.
Bresnan, J. 1972. *Theory of Complementation in English Syntax*. Ph.D. Dissertation, MIT.
Bybee, J. 1985. *Morphology. A study of the relation between meaning and form*. Amsterdam: Benjamins.
Cain, B. 2000. *Dhivehi (Maldivian): a synchronic and diachronic study*. Ph.D. Dissertation, Cornell University.
Cain, B. and J.W. Gair. 2000. *Dhivehi* (*Maldivian*). München: Lincom Europa.
Chapin, P. G. 1978. Easter Island. A Characteristic VSO Language. In *Syntactic Typology*, ed. W.P. Lehmann. 139–168 Sussex: The Harvester Press.
Chia, E. 1976. *Kom Tenses and Aspects*. Ph.D. Dissertation, Georgetown University.
Chiao, S.W. 1998. Complex Noun Phrases in Luo (available at: http://www.owlnet.rice.edu/~ling215/Luo/noun_phrases.html).
Cinque, G. 1994. On the evidence for partial N movement in the Romance DP. In *Paths Towards Universal Grammar. Studies in Honor of Richard S. Kayne*, eds. G. Cinque, J. Koster, J.-Y. Pollock, L. Rizzi & R. Zanuttini. 85–110. Washington (D.C.): Georgetown University Press.
Cinque, G. 1996. The antisymmetric programme: theoretical and typological implications. *Journal of Linguistics* 32: 447–464.
Cinque, G. 1999. *Adverbs and Functional Heads. A cross-linguistic perspective*. New York: Oxford University Press.
Cinque, G. 2002. "Complement and Adverbial PPs: Implications for Clause Structure", paper presented at the 25th Annual GLOW Colloquium, Amsterdam, April 9–11 2002 (published in Cinque 2006, pp.145–166).
Cinque, G. 2003. On Greenberg's Universal 20 and the Semitic DP. In *Grammar in Focus. Festschrift for Christer Platzack 18 November 2003.* vol. II, eds., L.-O. Delsing, C. Falk, G. Josefsson, and H. Sigurðsson. 243–25. Lund: Wallin and Dalholm.
Cinque, G. 2005a. Deriving Greenberg's Universal 20 and Its Exceptions. *Linguistic Inquiry* 36: 315–332.
Cinque, G. 2005b. A note on verb/object order and head/relative clause order. In *Grammar and Beyond. Essays in honour of Lars Hellan*, eds. M. Vulchanova and T.A. Åfarli, 69–89. Oslo: Novus Press.
Cinque, G. 2006. *Restructuring and Functional Heads. The Cartography of Syntactic Structures,* vol. 4. Oxford: Oxford University Press.
Cinque, G. 2007. Again on Tense, Aspect, Mood morpheme order and the "Mirror Principle". *University of Venice Working Papers in Linguistics*. vol. 17 (to appear in a Festschrift).
Cinque, G. (forthcoming) *The Syntax of Adjectives. A Comparative Study*.
Clifton, J.M. 1995. A Grammar Sketch of the Kaki Ae Language. *Work Papers of the Summer Institute of Linguistics* 39.33–80.
Coate, H.H.J. and L. Coate. 1970. *A Grammar of Ngarinjin, Western Australia.* Canberra: Australian Institute of Aboriginal Studies.
Colarusso, J. 1992. *A Grammar of the Kabardian Language*. Calgary: University of Calgary Press.
Cole, P., G. Hermon, and Y.N. Tjung. 2005. How irregular is WH in situ in Indonesian. *Studies in Language* 29: 553–581.
Collins, C. 2004. The Absence of the Linker in Double Object Constructions in N|uu. *Studies in African Linguistics* 33 (2):163–198.
Corbett, G. 1979. Adjective Movement. *Nottingham Linguistic Circular* 8: 1–10.
Croft, W. and E. Deligianni. 2001. Asymmetries in NP Word Order, ms., University of Manchester (available at: http://ling.man.ac.uk/Info/staff/WAC/Papers/NPorder.pdf).
Crowley, T. 2002. Port Sandwich. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley, 650–659. Richmond (Surrey): Curzon.
Cunha de Oliveira, C. 2003. Lexical Categories and the Status of Descriptives in Apinajé. *International Journal of American Linguistics* 69: 243–274.

Cutler, A., J.A. Hawkins, and G. Gilligan. 1988. The suffixing preference: a processing explanation. *Linguistics* 23: 723–758.

Dayley, J.P. 1989. *Tümpisa (Panamint) Shoshone Grammar*. Berkeley: University of California Press.

Douglas, W.H. 1981. Watjarri. In *Handbook of Australian Languages*. vol. 2, eds. R.M.W. Dixon and B.J. Blake. 196–272. Canberra: Australian National University Press.

Driem, G. van. 1987. *Grammar of Limbu*. Berlin: Mouton de Gruyter.

Dryer, M. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25: 123–195.

Dryer, M. 2000. *Word Order in Tibeto-Burman Languages*. Buffalo: State University of New York.

Dryer, M. 2001. Mon Khmer Word Order from a Crosslinguistic Perspective. In *Papers from the Sixth Annual Meeting of the Southeast Asian Linguistics Society. 1996*, eds. K. Adams and T.J. Hudak. 83–99. Tempe: Arizona State University (Program for Southeast Asian Studies. Monograph Series Press).

Dryer, Matthew S. 2007. *Word order*. In Language typology and syntactic description. Second Edition, Vol. I: Clause structure, ed. Timothy Shopen, 61131. Cambridge: Cambridge University Press.

Dutton, T.E. 1996. *Koiari*. München: Lincom Europa.

Efrat, B.S. 1969. *A Grammar of Non-Particles in Sooke, a Dialect of Straits Coast Salish*. Ph.D. Dissertation, University of Pennsylvania.

Elders S. 2000. *Grammaire Mundang*. Leiden: CNWS Publications.

Essegbey, J.A.B.K. 1993. The X-Bar Theory and the Ewe Noun Phrase. *University of Trondheim Working Papers in Linguistics* 19: 52–69.

Fassi Fehri, A. 1999. Arabic modifying adjectives and DP structures. *Studia Linguistica* 53: 105–154.

Feldpausch, T. and B. Feldpausch. 1992. Namia Grammar Essentials. In *Namia and Amanab Grammar Essentials*, ed. J.R. Roberts. Data Papers in Papua New Guinea Languages. vol. 39. 1–97. Ukarumpa via Lae (Papua New Guinea): Summer Institute of Linguistics.

Fernandez, F. 1967. *A Grammatical Sketch of Remo: A Munda Language*. Ph.D. Dissertation, University of North Carolina at Chapel Hill.

Foley, W. and R. van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.

Frantz, Donald G. 1991. *Blackfoot Grammar*. Toronto: University of Toronto Press.

Frajzyngier, Z. and E. Johnston (with A Edwards). 2005. *A Grammar of Mina*. Berlin: Mouton De Gruyter.

Gair W.J. and J.C. Paolillo. 1997. *Sinhala*. München: Lincom Europa.

Genetti, C. and K. Hildebrandt. 2004. The Two Adjective Classes in Manange. In *Adjective Classes. A Cross-linguistic Typology*, eds. R.M.W. Dixon and A.Y. Aikhenvald. 74–96. Oxford: Oxford University Press.

Gerdts, Donna. 1982. *Object and Absolutive in Halkomelem Salish*. Ph.D. Dissertation, University of California at San Diego.

Geytenbeek, B. and H. Geytenbeek. 1971. *Gidabal Grammar and Dictionary*. Canberra: Australian Institute of Aboriginal Studies.

Giridhar, P.P. 1994. *Mao Naga Grammar*. Manasagangotri: Central Institute of Indian Languages.

Giusti, G. 1996. Is there a FocusP and a TopicP in the Noun Phrase Structure? *University of Venice Working Papers in Linguistics* 6 (2):105–128.

Ghosh, A. 1994. *Santali. A Look into Santal Morphology*. New Delhi: Gyan Publishing House.

Glass, A. and D. Hackett. 1970. *Pitjantjatjara Grammar. A Tagmemic View of the Ngaayatjara (Warburton Ranges) Dialect*. Canberra: Australian Institute of Aboriginal Studies.

Greenberg, J. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed., J. Greenberg, 73–113. Cambridge (Mass.): MIT Press.

Haddican, B. 2002. Aspects of DP Word Order Across Creoles", paper presented at the CUNY/SUNY/NYU Linguistics Mini-Conference, April 20, 2002 (abstract available at: http://www.nyu.edu/gsas/dept/lingu/events/minico02/abstract/haddican.pdf).

Hamel, P.J. 1994. *A Grammar and Lexicon of Loniu, Papua New Guinea.* Canberra: Australian National University (Pacific Linguistics, Series C-103).
Hansson, I.-L. 2003. Akha. In *The Sino-Tibetan Languages*, eds., G. Thurgood and R. LaPolla, 236–251. London: Routledge.
Harriehausen, B. 1990. *Hmong Njua*. Tübingen: Niemeyer.
Harris, H.R. II. 1977. *A Grammatical Sketch of Comox*. Ph.D. Dissertation, University of Kansas.
Harris, K. 1990. Nend Grammar Essentials. In *Two Grammatical Studies*. Data Papers in Papua New Guinea Languages. vol. 37, eds. J.R. Roberts. 73–156. Ukarumpa via Lae (Papua New Guinea): Summer Institute of Linguistics.
Harms, P. L. 1994. *Epena Pedee Syntax* (Studies in the Languages of Colombia, 4). Summer Institute of Linguistics, and University of Texas at Arlington.
Hawkins, J.A. 1983. *Word Order Universals*. New York: Academic Press.
Hawkins, J.A. 1988. On explaining some left-right asymmetries in syntactic and morphological universals. In *Studies in Syntactic Typology*, eds. M. Hammond, E.A. Moravcsik, and J.R. Wirth, 321–357. Amsterdam: Benjamins.
Hawkins, J.A. 1998. Adjacency, linear precedence and the theory of dependency strength", ms., Los Angeles: University of Southern California.
Hawkins, J.A. and A. Cutler 1988. Psycholinguistic Factors in Morphological Asymmetry. In *Explaining Language Universals*, ed. J.A. Hawkins, 280–317. Oxford: Blackwell.
Hawkins, J.A. and G. Gilligan. 1988. Prefixing and suffixing universals in relation to basic word order. In *Papers in Universal Grammar: Generative and Typological Approaches*, eds. J.A. Hawkins and H.K. Holmback. (Lingua Special Issue 74: 219–259).
Hayward, D. 1984. *The Arbore Language: A First Investigation Including a Vocabulary*. Hamburg: Buske.
Heath, J. 2005. *Tondi Songway Kiini (Songhay, Mali). Reference Grammar and TSK-English-French Dictionary*. Stanford: CSLI Publications.
Heine, B. 1980. Elmolo. In *The Non-Bantu Languages of Kenya*, ed. B. Heine, 173–218. Berlin: Reimer.
Heine, B. 1981. Determination in some East African Languages. In *Wege zur Universalienforschung: sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler*, eds. G. Brettschneider and C. Lehmann 180–186. Tübingen: Narr.
Helmbrecht, J. 2004. Are there adjectives in Hocank (Winnebago)? *Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt*. Nr.14.
Hershberger, H. 1964. Gugu-Yalanji Noun Phrases. In *Papers on the Languages of the Australian Aborigines*, eds. R. Pittman and H. Kerr. 83–90. Canberra: Australian Institute of Aboriginal Studies.
Hetzron, Robert. 1978. On the Relative Order of Adjectives. In *Language Universals*, ed. H. Seiler, 165–184. Tübingen: Narr.
Hinterhölzl, Roland. 2002. Event-related Adjuncts and the OV/VO Distinction. In *Proceedings of the 20th West Coast Conference in Formal Linguistics*, eds. K. Magerdoomian and L.A. Bar-el, 276–289. Somerville: Cascadilla.
Hogan, D. 1999. *Urak Lawoi'*. München: Lincom Europa.
Holt, D. 1999. *Pech (Paya).* München: Lincom Europa.
Hyman, Larry. M. ed., 1979. *Aghem Grammatical Structure*. Southern California Occasional Papers in Linguistics, no.7. Los Angeles: University of Southern California.
Hyman, Larry. M. 1981. *Noni Grammatical Structure*. Southern California Occasional Papers in Linguistics, no.9. Los Angeles: University of Southern California.
Jelinek, E. and R. Demers. 1997. Reduplication as a Quantifier in Salish. *International Journal of American Linguistics* 63: 302–315.
Jones, R.B. 1970. Classifier constructions in southeast Asia. *Journal of the American Oriental Society* 90: 1–12.
Kayne, R. 1994. *The Antisymmetry of Syntax.* Cambridge (Mass.): MIT Press.
Kölver, U. 1978. On Newari noun phrases. In *Language Universals*, ed. H. Seiler. 277–300.Tübingen: Narr.

Koopman, H. 2003. On the Parallelism of DPs and Clauses: Evidence from Kisongo Maasai, ms., UCLA.
Koopman, H. and A. Szabolcsi. 2000. *Verbal Complexes*. Cambridge (Mass.): MIT Press.
Koshal, S. 1979. *Ladakhi Grammar*. Delhi: Motilal Banarsidass (B.G. Misra ed.).
Kroeker, M. 2001. A Descriptive Grammar of Nambikuara. *International Journal of American Linguistics* 67: 1–87.
Langdon, M. 1970. *A Grammar of Diegueño. The Mesa Grande Dialect*. Berkeley: University of California Press.
LaPolla, R. J. 2003. Dulong. In *The Sino-Tibetan Languages*, eds G. Thurgood and R. LaPolla. 674–682. London: Routledge.
Lawton, R. 1993. *Topics in the Description of Kiriwina*. Canberra: Australian National University (Pacific Linguistics, Series D-84).
Lee, R. 1994. *Madak Noun Phrases. Madak Sentences*. Ukarumpa (Papua New Guinea): Summer Institute of Linguistics.
Lu, B. 1998. *Left-Right Asymmetries of Word Order Variation. A Functional Explanation*. Ph.D. Dissertation. University of Southern California.
Lu, B. n.d. Ordering of Time, Duration, Location, Instrument, Manner, ms., University of Southern California.
Lux, D. and C. Lux 1996. Tone in the Noòni Noun Phrase: orthographic considerations, ms., Yaoundé: Summer Institute of Linguistics (available on-line in the Summer Institute of Linguistics, Cameroon, website).
Lynch, J. 1998. *Pacific Languages. An Introduction.* Honolulu: University of Hawai'i Press.
Lynch, J. 2002. Xârâcùù. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 765–775. Richmond (Surrey): Curzon.
Martin, J.B. 2000. Creek Voice: Beyond Valency. In *Changing Valency. Case Studies in Transitivity*, eds. R.M.W. Dixon and A.Y. Aikhenvald. 375–403. Cambridge: Cambridge University Press.
Matthewson, L. (2003) "On the methodology of semantic fieldwork", in *Issues in Language and Linguistics. Current Research at UBC*, eds. C. Gillon, C. Ravinski, and R. Wojdak. *University of British Columbia Working Papers in Linguistics* 13.37–78.
Mazaudon, M. 2003. Tamang. In *The Sino-Tibetan Languages*, eds. G. Thurgood and R. LaPolla. 291–314. London: Routledge.
Minch, A.S. 1991. *Essential Elements of Amanab Grammar*. M.A. Dissertation, University of Texas at Arlington.
Mous, M. 1997. The Position of the Object in Tunen. In *Object Positions in Benue-Kwa*, eds. R.-M. Déchaine and V. Manfredi. 123–137. The Hague: Holland Academic Graphics.
Moyse-Faurie, C. 1995. *Le Xârâcùù. Langue de Thio-Canala (Nouvelle-Calédonie). Eléments de syntaxe*. Paris: Peeters.
Mulder, J.G. 1994. *Ergativity in Coast Tsimshian (Sm'algyax)*. Berkeley: University of California Press.
Nagano, Y. 2003. Cogtse Gyarong. In *The Sino-Tibetan Languages*, eds. G. Thurgood and R.J. LaPolla. 469–489. London: Routledge.
Nagaraja, K.S. 1985. *Khasi. A Descriptive Analysis*. Pune: Deccan College.
Nguyen, T. H. 2004. *The Structure of the Vietnamese Noun Phrase*. Ph.D. Dissertation, Boston University.
Nilsen, Ø. and N. Vinokurova. 2000. Generalized verb raisers. In *Proceedings of the 2000 International Workshop on Generative Grammar*, eds. Y.J. Jang and J.-S. Kim. Seoul: Hansung University.
Noonan, M. 1992. *A Grammar of Lango*. Berlin: Mouton de Gruyter.
Oates, L.F. 1964. *A Tentative Description of the Gunwinggu Language* [*of Western Arnhem Land*]. Oceania Linguistic Monograph No.10, University of Sydney.
O'Connor, M. 1992. *Topics in Northern Pomo Grammar*. New York: Garland.
Opgenort J. R. 2005. *A Grammar of Jero*. Leiden: Brill.

Osborne, C.R. 1974. *The Tiwi Language*. Canberra: Australian Institute of Aboriginal Studies.
Osumi, M. 1995. *Tinrin Grammar*. Honolulu: University of Hawai'i Press.
Pawley, A. and T. Sayaba. 1982. A sketch grammar of the Nabukelevu language of Kadavu. *Te Reo. Journal of the Linguistic Society of New Zealand* 25: 35–93.
Pearce, E. 2002. Phrasal Movement within the Mori DP, ms., Victoria University of Wellington.
Pearson, M. 2000. Two types of VO languages. In *The Derivation of VO and OV*, ed. P. Svenonius. 327–363. Amsterdam: Benjamins.
Plank, F. 2006. Canonical and non-canonical order in noun phrases (and where is information structure?). Paper presented at the workshop *DP-internal Information Structure: Topic, Focus and other Illocutionary forces*. University of Utrecht, 17–18 November 2006.
Quesada, D. J. 1999. Chibchan, with special reference to participant highlighting. *Linguistic Typology* 3/2: 209–258.
Rakowski, A. and L. Travis. 2000. V-initial Languages: X or XP Movement and Adverbial Placement. In *The Syntax of Verb Initial Languages*, eds. A. Carnie and E. Guilfoyle. 117–141. New York: Oxford University Press.
Ramirez, C. 1998. The Kenyang Noun Phrase, ms., Yaoundé, Summer Institute of Linguistics (available on-line at the Summer Institute of Linguistics, Cameroon, website).
Renck, G.L. 1975. *A Grammar of Yagaria*. Canberra: Australian National University (Pacific Linguistics, B–40).
Rice, K. 1989. *A Grammar of Slave (Dene)*. Berlin: Mouton de Gruyter.
Rijkhoff, J. 1990. Explaining word order in the noun phrase. *Linguistics* 28: 5–42.
Rijkhoff, J. 1998. Order in the noun phrase of the languages of Europe. In *Constituent Order in the Languages of Europe*, ed. A. Siewierska. 321–382. Berlin: Mouton de Gruyter.
Rijkhoff, J. 2002. *The Noun Phrase*. Oxford: Oxford University Press.
Ross, J.R. 1967.*Constraints on Variables in Syntax*. Ph.D. Dissertation, MIT.
Ross, M. 2002a. Kele. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 123–147. Richmond (Surrey): Curzon.
Ross, M. 2002b. Kairiru. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 204–215. Richmond (Surrey): Curzon.
Ross, M. 2002c. 'Ala'ala. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 347–361. Richmond (Surrey): Curzon [adapted from Symonds n.d. 1989].
Ross, M. 2002d. Kaulong. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 387–409. Richmond (Surrey): Curzon.
Ross, M. 2002e. Sisiqa. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 459–466. Richmond (Surrey): Curzon.
Rouveret, A. 1994. *Syntaxe du gallois*. Paris: CNRS Editions.
Saul, J.E. and N. Freiberger Wilson. 1980. *Nung Grammar*. Summer Institute of Linguistics and University of Texas at Arlington.
Schuh, R.G. 1998. *A Grammar of Miya*. Berkeley: University of California Press.
Schweikert, W. 2005. *The Order of Prepositional Phrases in the Structure of the Clause*. Amsterdam: Benjamins.
Scott, G.-J. 2002. Stacked adjectival modification and the structure of the nominal phrases. In *Functional Structure in DP and IP. The Cartography of Syntactic Structures,* vol. 1, ed. G. Cinque. 91–120. New York: Oxford University Press.
Serzisko, F. 1989. The Kuliak Languages: a structural comparison. In *Topics in Nilo-Saharan Linguistics*, ed. M.L. Bender. 385–404. Hamburg: Buske.
Shaul, D.L. 1986. *Topics in Nevome Syntax*. Berkeley: University of California Press.
Siewierska, A. and L. Uhlíová. 1997. An overview of word order in Slavic languages. In *Constituent Order in the Languages of Europe*, ed. A. Siewierska. 105–149. Berlin: Mouton de Gruyter.
Sigurðsson, H. Á. 1993. The structure of the Icelandic NP. *Studia Linguistica* 47: 177–197.
Smith, I. and S. Johnson. 2000. Kugu Nganhcara. In *Handbook of Australian Languages.* vol. 5, eds. R.M.W. Dixon and B.J. Blake. 357–489. Melbourne: Oxford University Press.

Soukka, M. 2000. *A Descriptive Grammar of Noon*. München: Lincom Europa.
Sproat, R. and C. Shih. 1991. The Cross-Linguistics Distribution of Adjectival Ordering Restrictions. In *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda*, eds. C. Georgopoulos and R. Ishihara. 565–593. Dordrecht: Kluwer.
Staalsen, P. 1965. *Iatmul Grammar Sketch* (http://www.sil.org/pacific/png/pubs/0000541/Iatmul_Gram_Sketch.pdf).
Staalsen, P. 1972. Clause Relationships in Iatmul. *Papers in New Guinea Linguistics* 15: 45–68 (Pacific Linguistics, Series A-31).
Steele, S. (with A. Akmajian, R. Demers, E. Jelinek, C. Kitagawa, R. Oehrle, and T. Wasow). 1981. *An Encyclopedia of AUX. A Study of Cross-Linguistic Equivalence*. Cambridge (Mass.): MIT Press.
Svenonius, P. 2006. 1… 3–2, ms., University of Tromsø (available on line at http://ling.auf.net/lingBuzz/000200).
Sy, M. 2003. On Wolof Adjectival Constructions, unpublished paper, UCLA.
Terzi, A. 1999. Clitic Combinations, their Hosts and their Ordering. *Natural Language and Linguistic Theory* 17: 85–121.
Thompson, S.A., J.S.-Y. Park, C.N. Li 2006. *A Reference Grammar of Wappo*. Berkeley: University of California Press.
Thornell, C. 1997. *The Sango Language and Its Lexicon* (Travaux de l'Institut de Linguistique de Lund 32). Lund: Lund University Press.
Tourneux, H. 2003. L'encodage de la localization, de la direction et du movement dans les langues "kotoko" du cameroun. In *Motion, Direction and Location in Languages. In Honor of Zygmunt Frajzyngier*, eds. E. Shay and U. Seibert. 287–297. Amsterdam: Benjamins.
Tryon, D.T. 1974. *Daly Family Languages, Australia*. Canberra: Australian National University.
Tryon, D.T. 2002. Buma. In *The Oceanic Languages*, eds. J. Lynch, M. Ross, and T. Crowley. 573–586. Richmond (Surrey): Curzon.
Valenzuela, Pilar. 2002. *Relativization in Shipibo-Konibo: A typologically-oriented study.* München: Lincom Europa.
Vangsnes, Øystein. 2004. Rolling up the Scandinavian noun phrase, ms., University of Tromsø.
Wiersma, G. 2003. Yunnan Bai. In *The Sino-Tibetan Languages*, eds. G. Thurgood and R. LaPolla, 651–673. London: Routledge.
Wilkendorf, P. n.d. *Sketch Grammar of Nomaándé: Sections 1–4*. Yaoundé: Summer Institute of Linguistics (available on line in the Summer Institute of Linguistics, Cameroon, web site).
Willis, David. 2006. Against N-raising and NP-raising analyses of Welsh noun phrases.*Lingua* 116: 1807–1839.
Wurmbrand, Suzi. 2004. West Germanic verb clusters: The empirical domain. In *Verb Clusters: A Study of Hungarian, German, and Dutch*, eds. K..É. Kiss and H. van Riemsdijk. 43–85. Amsterdam: Benjamins.
Zegrean, I.-G. 2007. *Towards Source of Motion in Romanian*. M.A. Dissertation, University of Venice.

The Branching Direction Theory of Word Order Correlations Revisited[1]

Matthew S. Dryer

Abstract In this paper, I discuss implications for the Branching Direction Theory proposed by Dryer (1992a) for accounting for the Greenbergian word order correlations if one assumes much flatter constituent structures than those assumed in Dryer (1992a) and in much work in generative syntax. Some correlations that are not accounted for by the Branching Direction Theory if we assume flatter constituent structures, such as the order of article and noun, can be accounted for by proposals in Hawkins (1994).

Keywords Word order · word order correlations · sentence processing · branching direction

In Dryer (1992a), I presented evidence from a sample of 625 languages on which pairs of elements correlate with the order of object and verb, and which do not (Greenberg 1963). I argued that the word order correlations reflect a tendency for languages to be consistently left-branching or consistently right-branching, what I referred to as the Branching Direction Theory (the BDT), and proposed that this tendency reflects processing difficulties associated with mixing left- and right-branching. The predictions of the BDT depend heavily, however, on one's assumptions about constituent structure. A number of the correlations require assuming fairly hierarchical constituent structures, and are not predicted by the BDT if one assumes flatter constituent structures. In this paper, I discuss a number of these correlations, arguing that some of them can be explained by a combination of processing considerations and other principles, while a few remain unexplained under assumptions of flat constituent structures.

M.S. Dryer
University at Buffalo

[1] Part of the research for this paper was made possible by Social Sciences and Humanities Research Council of Canada Grants 410-810949, 410-830354, and 410-850540, by National Science Foundation Research Grant BNS-9011190, and by support from the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. I am indebted to Lea Brown, Jack Hawkins, Bill Croft, an anonymous reviewer, and members of the audience at the conference in Bologna for comments on earlier drafts of this paper.

S. Scalise et al. (eds.), *Universals of Language Today*, 185–207,
DOI: 10.1007/978-1-4020-8825-4_10,

1 What Correlates with the Order of Object and Verb?

While my current database is considerably larger than it was at the time I wrote Dryer (1992a), now containing at least partial data for over 1500 languages, the additional languages do not change the evidence presented in Dryer (1992a) as to which pairs of elements correlate with the order of object and verb. In this section, I summarize the conclusions regarding this.

In order to discuss the correlations, it is useful to have a way of referring to the various pairs of elements that correlate with the order of object and verb and the members of each pair. To do this, I will say that if a pair of elements X and Y is such that X tends to precede Y significantly more often in VO languages than in OV languages, then <X, Y> is a *correlation pair*, and X is a *verb patterner* and Y an *object patterner* with respect to this pair. In Dryer (1992a) I provided evidence that each of the pairs of elements in Table 1 is a correlation pair (except for the order of complementizer and clause, for which data is given on p. 199 below).

Table 1 Correlation Pairs

Verb Patterner	Object Patterner	Example
verb	object	*ate + the sandwich*
adposition	NP	*on + the table*
noun	relative clause	*movies + that we saw*
article	N|	*the + tall man*
copula verb	predicate	*is + a teacher*
'want'	VP	*wants + to see Mary*
tense-aspect auxiliary verb	VP	*has + eaten dinner*
negative auxiliary	VP	
complementizer	S	*that + John is sick*
question particle	S	
adverbial subordinator	S	*because + Bob has left*
plural word	N|	
noun	genitive	*father + of John*
adjective	standard of comparison	*taller + than Bob*
verb	PP	*slept + on the floor*
verb	manner adverb	*ran + slowly*

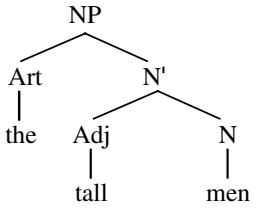
2 Explaining the Word Order Correlations

In Dryer (1992a), I argued in detail that the word order correlations cannot be handled in terms of consistent ordering of heads and dependents. Such a theory would predict that adjectives, demonstratives, and numerals ought to be object patterners, but they are not; the order of these three elements with respect to the noun does not correlate with the order of object and verb (Dryer 2005a). Nor can the correlations be handled in terms of consistent ordering of heads and complements. Such a theory would fail to predict that the order of relative clause and noun and the order of adpositional phrase and verb do correlate with the order of object and verb.

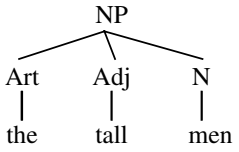
I proposed instead what I called the Branching Direction Theory (BDT), according to which verb patterners are nonphrasal while object patterners are phrasal, with the effect that languages tend towards being either consistently left-branching or consistently right-branching. In addition, I proposed that structures with consistent left-branching or consistent right-branching are easier to process than structures that involve a mixture of left- and right-branching.

A practical problem with the BDT is that it depends on one's assumptions regarding constituent structure. For example, in order to account for the fact that articles are verb patterners, the BDT requires that we assume a structure like that in (1a), where the article combines with a phrasal category like an N'. If, on the other hand, we assume a flatter constituent structure like that in (1b), then the BDT would fail to predict that articles are verb patterners, since they would not be combining with a phrasal category.

(1) a.

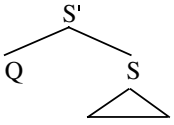


b.

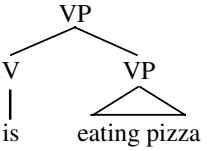


Similarly, unless we assume that polar question particles combine with clauses, as in (2a), rather than simply being one constituent of the clause, the BDT fails to predict that polar question particles are verb patterners; and unless we assume that auxiliary verbs combine with VPs, as in (2b), the BDT fails to predict that auxiliary verbs are verb patterners.

(2) a.



b.



3 Flat Constituent Structure

Because the BDT depends on assumptions about constituent structure that not everyone believes, the BDT will be unconvincing to anyone whose assumptions about constituent structure are different from those required for the BDT to make the correct predictions, and more specifically, will be unconvincing to anyone who assumes that constituent structure is flatter than what is required for the BDT to work. Now perhaps this would not worry me particularly if the constituent structures required by the BDT were ones that I myself believed in. The problem, however, is that over the past fifteen years, my own views about constituent structure have changed, so

that I now am one of those who believe that constituent structure is flatter than has often been assumed in generative grammar over the past thirty-five years. What this means is that given my own assumptions about constituent structure, the BDT fails to account for a number of the word order correlations.

The discussion here of why one might adopt flat constituent structures will not do the topic justice, but for reasons of space, my comments are necessarily brief. The idea that constituent structure is flatter than is often assumed in generative grammar is a view that is explicit or implicit in a number of approaches. For example, the idea that articles do not combine with N$^{|}$ s is implicit in much work from the early days of generative grammar, where the trees assumed by generative linguists often involved a flatter structure along the lines of (1b). More recently, Culicover and Jackendoff (2005) argue for flatter constituent structures. Furthermore, flatter constituent structures are implicit in most work within Basic Linguistic Theory (Dixon 1997), the theoretical framework assumed in most descriptive grammars written within the last twenty years and implicit in much work in linguistic typology. Such descriptive grammars all assume that there are noun phrases and clauses, which means that they are implicitly assuming that these at least are constituents, but it is relatively rare that such descriptions will assume anything analogous to an N$^{|}$, and it is if anything even rarer for someone to describe a polar question particle as combining with a clause. In discussing the structure of noun phrases, most such grammars treat articles simply as one type of modifier of the noun. Similarly, most such grammars treat question particles as simply one constituent within the main clause.

The fact that clauses and noun phrases seem to be universally recognized as constituents and have been since the time of traditional grammar reflects a property of the notion of constituent that often seems to be forgotten by linguists of very different persuasions. And that is that constituent structure is largely semantic. The reason that people have always recognized clauses and noun phrases as constituents is that they are clearly semantic units, the clause corresponding to a proposition, or a situation, or an event, or a state of affairs, the noun phrase corresponding to things. Students in introductory linguistics classes are generally better at identifying clauses and noun phrases as a constituents, not because they have natural talents as linguists, but because, in identifying constituents, they are primarily tapping in on their knowledge of the meaning of the sentence and their awareness of which words go together to form semantic units.

The strongest claim one might make, in fact, is that constituent structure is entirely semantic, that constituent structure trees simply represent one aspect of the meaning of the sentence, of which words go together to form semantic units. I think that in its strongest form, such a claim will not work. For example, the two sentences in (3) have the same (surface) constituent structure, but arguably different semantic structures.

(3) a. John is likely to win the election.
    b. John is eager to win the election.

There is a clear sense in which the syntactic structure of these two sentences is closer to the semantic structure of (3b), and (3a) in some sense has a syntactic constituent

structure that is somewhat different from its semantic structure. Similarly, the two structures in (4) are two ways of expressing the same meaning across different languages, but have different constituent structures.

(4) a. John caused [Bill to fall].
b. John [caused-fall] Bill.

The different syntactic structures here do not involve a difference in what is a semantic unit.

While examples like those in (3) and (4) argue against a strong claim that syntactic structure is purely semantic, this does not alter the fact that these examples are exceptional. To a large extent constituent structure is semantic, much like the fact that membership in word classes is usually largely predictable from the meaning of words, but not entirely. The extent to which constituent structure is semantic has also been obscured by the sorts of constituent structures that have been popular in generative grammar over the past thirty-five years, first in adopting more hierarchical constituent structures which claim that a lot of sequences of words are constituents that were not viewed as such before, with the added effect that many of the new constituents are less clearly semantic units, and second by assuming constituent structures that are more abstract relative to surface structure. If one assumes flatter constituent structures, then the constituents in such flat constituent structures are more likely to be semantic units.

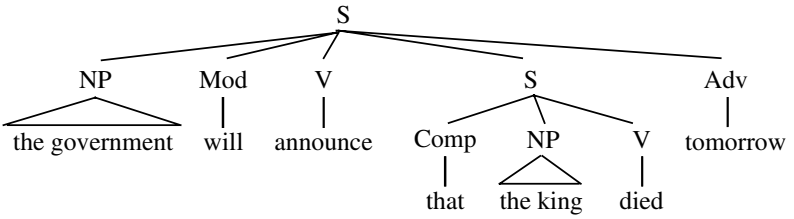
But the notion of semantic units or semantic constituents is also of specific relevance to the BDT. Underlying that theory is the idea that the word order correlations reflect parsing or processing difficulties associated with certain sorts of syntactic structures. But ultimately, the reason that people parse sentences is to understand the meaning of the sentence. There is often a tendency to view parsing as a process of assigning syntactic structures to sentences, but the final result is the hearer's assigning a meaning to the sentence. Parsing is ultimately a matter of determining the semantic units or semantic constituents of a sentence. In fact, one possible view of parsing is that that is all it is: a process of determining what the semantic units of a sentence are, of determining which words go together semantically. Thus whether or not syntactic constituent structure is flat or hierarchical may not really matter, if what really matters is the semantic units. In other words, one might claim that syntactic structure is not flat, but that when people parse sentences, they only try to identify syntactic constituents that are also semantic units. And while I will formulate the rest of this paper in terms of flat syntactic structures, I could equally well have formulated it in terms of flat semantic structures, while remaining neutral on the question of whether syntactic structures are also flat.

To summarize what I have said so far, parsing sentences correctly means assigning sentences the right meaning. Structures that are difficult to parse are ones that present difficulty for hearers to assign the correct meaning to. In other words, structures that are difficult to parse are ones for which speakers have difficulty identifying the semantic constituents. Let me illustrate the point with a few examples. Consider the sentence in (5).

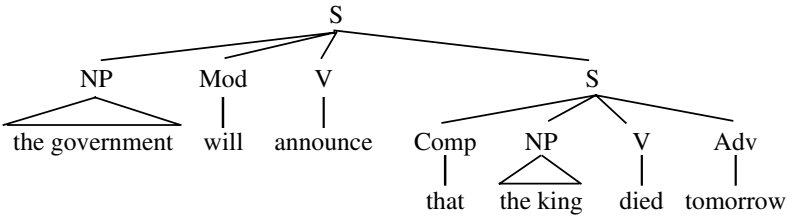
(5) The government will announce that the king died tomorrow.

From a purely syntactic point of view, this sentence has two possible structures, given in (6) and (7). Note that the trees I give are flatter than what is customary nowadays, though they are not radically different from the trees used by generative linguists in the 1960s, except my trees do not recognize a VP constituent (cf. Van Valin and LaPolla 1997: 217–218). But the flatness of these trees, while reflecting the structures I will assume below, is actually irrelevant to the point I am currently making.

(6)



(7)

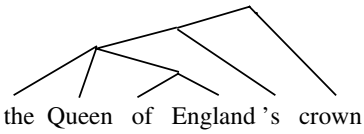


This sentence is interesting in that there is a tension between the automatic strategies of the human parser and what sort of meaning makes sense. The nature of the human parser is such that it tries to assign the sentence the structure in (7). This preference has been expressed in many different ways, one of them being the principle of Late Closure of Frazier (1978). But the meaning associated with (7) doesn't make sense: the past tense of the verb *died* is incompatible with the meaning of the adverb *tomorrow*. The human parser is such that people may never realize that the sentence has another meaning, that corresponding to (6).

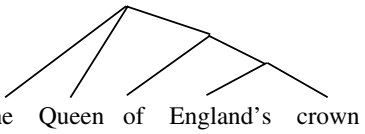
But the main point I want to make about this example is that if someone assigns the sentence the wrong constituent structure, that means that they have assigned it the wrong semantic structure: the syntactic difference between (6) and (7) is equally well a semantic difference, and the difference in the syntactic structures directly represents the difference in meaning: in (7), *the king died tomorrow* is a semantic unit, while in (6) it is not. In other words, the processing difficulty associated with (6) can be described as a difficulty assigning it the right syntactic structure, but that difficulty in assigning it the right syntactic structure is equivalent to the difficulty assigning the sentence the right meaning. If one doesn't assign the sentence the constituent structure of (6), one doesn't understand the meaning of the sentence.

Or consider the pair of constituent structures in (8). Both are possible syntactic structures, but again one of them, namely (8b), is semantically anomalous. If on hearing a sentence containing this phrase, someone assigns it the syntactic structure in (8b), or, equivalently, assigns it the semantic structure in (8b), that means that they have not understood the sentence.

(8) a.

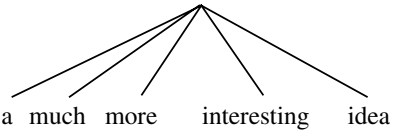


b.

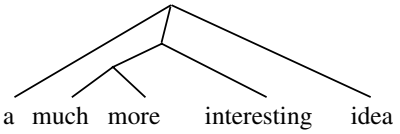


And if one were to assign the syntactic or semantic structure in (9a) to the phrase *a much more interesting idea*, that would mean that one had not understood the phrase, since understanding the phrase requires that one recognize that it has a semantic structure like that in (9b).

(9) a.

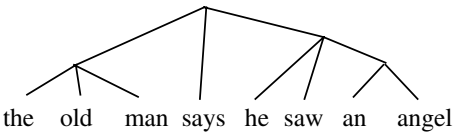


b.



And interpreting the sentence in (10) requires that one realize that *the old man* belongs to the clause with the verb *says*, but that *he* and *an angel* belong to the clause with the verb *saw*. If one doesn't realize this, one has not understood the sentence.

(10)



The general point is that when I talk about processing difficulties associated with certain constituent structures, I am talking about difficulties in communication, difficulties in assigning the correct meaning to a sentence.

This brings us back to the issue of flat constituent structure and the BDT. Structures that are difficult to parse are ones where hearers have difficulty recognizing which words go together semantically. But that means that the constituents that are crucial are those that are clearly semantic units, like clauses and noun phrases. Hence for the BDT to be a convincing explanation for the word order correlations, it must work for flat structures, structures that represent the semantic units. If the BDT depends on structures that are irrelevant to meaning, then it fails as an adequate account.

So which of the correlation pairs in Table 1 above involve pairs of elements where the object patterner is clearly a semantic unit? Admittedly, there is room for considerable disagreement as to what constitutes a semantic unit, so it may not be clear in all cases whether the object patterner is a semantic unit. However, the pairs in Table 2 are pairs for which I believe a good case can be made that the object patterners are semantic units, either because they involve a clause (including VPs without a subject, which can be analysed as clauses without a subject) or a noun phrase (possibly including an adposition).

On the other hand, the pairs in Table 3 are ones for which it is less clear that the object patterners are semantic units.

What I have given as object patterners in Table 3 are as I listed them in Table 1 above (and in Dryer 1992a). But under the assumption that these object patterners are not semantic units, i.e., that the verb patterners are sisters to the constituents of the object patterners, these correlations would need to be reformulated. In fact, if these verb patterners occur as the initial (or final) constituent amongst a number of sister constituents, then the very notions of object patterner and correlation pair become problematic. At most, we can say that the verb patterners tend to occur first within their mother constituent more often in VO languages than in OV languages. I will therefore formulate the discussion below in terms of the position of the verb patterners in Table 3.

The question therefore is whether we can find alternative explanations for why the verb patterners in Table 3 are verb patterners. For reasons of space, I will only be able to discuss three of these verb patterners, namely articles, complementizers, and auxiliary verbs.

Table 2 Correlation pairs where the object patterner is a semantic unit

Verb Patterner	Object Patterner	Example
verb	object	*ate* + *the sandwich*
copula verb	predicate	*is* + *a teacher*
'want'	VP	*wants* + *to see Mary*
noun	genitive	*father* + *of John*
noun	relative clause	*movies* + *that we saw*
adjective	standard of comparison	*taller* + *than Bob*
verb	PP	*slept* + *on the floor*

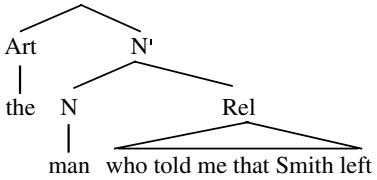
Table 3 Correlation pairs where is it less clear that the object patterner involves a semantic unit

Verb Patterner	Object Patterner	Example
article	N	*the* + *tall man*
plural word	N	
adposition	NP	*on* + *the table*
tense-aspect auxiliary verb	VP	*has* + *eaten dinner*
negative auxiliary	VP	
complementizer	S	*that* + *John is sick*
adverbial subordinator	S	*because* + *Bob has left*
question particle	S	

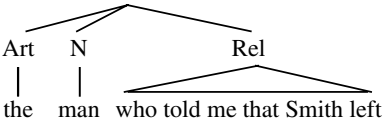
4 The Position of Articles

The first verb patterner I will discuss is articles. As mentioned above, for the BDT to work, one must assume a structure in which the article combines with an N', as in (11a) (or an analysis in which the article is a determiner and combines with an NP to form a DP); if, on the other hand, we assume a flatter structure, as in (11b), the article would not be combining with a semantic unit.

(11) a.



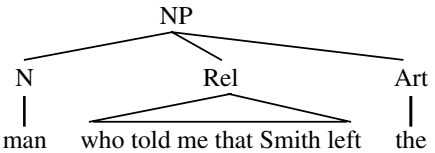
b.



In the particular example in (11b), one of the two constituents that the article is combining with is a relative clause, which is a phrasal category. However, it is not in general the case that articles combine with phrasal categories, since they may combine with just a noun, or just a noun plus one or more nonphrasal modifiers of the noun, like numerals or adjectives. In other words, if we assume a structure like that in (11b), articles no more combine with phrasal categories than adjectives or numerals do, but the latter modifiers are not verb patterners. In short, the BDT fails to account for the fact that articles are verb patterners while adjectives and numerals are not, if we assume flat structures like that in (11b).

So let us search for an alternative explanation for why articles might be verb patterners. In other words, what we need to do is explain why VO languages are more likely than OV languages to employ the word order in (11b) than the word order in (12), where the article occurs at the end of the noun phrase.

(12)



It turns out that there is a proposal in the literature that appears to explain this. Hawkins (1994, 2004) proposes a principle like that in (13), which I will refer to as Hawkins' Principle. Hawkins' theory is more complex than I will portray it in this paper, and my discussion will simplify it in some ways, though I believe that this simplification does not affect the argument.

(13) Hawkins' Principle
Structures with shorter constituent recognition domains (CRDs) are easier to process.

(Hawkins 1994, 2004)

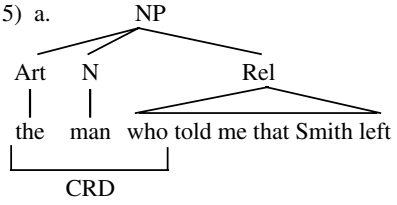
The notion of *constituent recognition domain* (CRD) is central to Hawkins' theory. Again simplifying things somewhat, the CRD of a constituent is the stretch of words in that constituent starting with the first word that allows the parser to construct the first daughter of that constituent and ending with the first word that allows the parser to construct the last daughter of that constituent, where a word allows the parser to construct a phrase if the parser can infer the category of the daughter phrase from that word. Hawkins calls words that allow the parser to construct phrases *mother-node constructing categories* (MNCCs). The clearest instances of MNCCs are heads. For example, if the parser encounters a noun, it can normally infer that the noun is part of a noun phrase, and the noun therefore allows the parser to construct a noun phrase. In addition to heads, various other words also serve as MNCCs. For example, articles are MNCCs for noun phrases; i.e. if the parser encounters an article, it can thereby infer that there is a noun phrase. If the first daughter of a constituent consists of one word, the CRD will start with that word. Similarly, if the last daughter of a constituent consists of one word, the CRD will end with that word.

Hawkins' theory is similar in spirit to the BDT in that it predicts that certain sorts of structures are less common in language because they are more difficult to process. Strictly speaking, the BDT is just an hypothesis about what provides a general categorization of verb patterners and object patterners; the proposal that the tendency to avoid mixing left and right branching is motivated by processing difficulties associated with such structures is actually a separate hypothesis. Hawkins' Principle furthermore makes similar predictions to the BDT. It predicts, for example, that structures that involve a mixture of left and right branching will be less common in language because they in general have longer CRDs (see, for example, Hawkins 1994: 96). On the other hand, it is a broader theory in a number of ways, for example in that it predicts lower frequency of structures that are independent of the word order correlations. It predicts, for example, that (14a) is preferred over (14b) because the CRD for the VP (or the clause if one assumes a flat structure) is shorter in (14a).

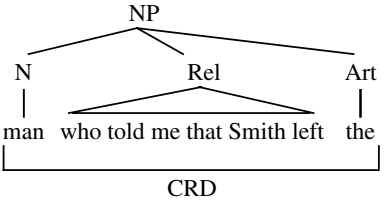
(14) a. He brought over the books that I asked for.
b. He brought the books that I asked for over.

But most important for this paper is the fact that Hawkins' Principle appears to provide at least a partial explanation for some of the correlations that the BDT fails to account for under assumptions of flat structures. For example, consider the structure in (11b) and (12) above, repeated here as (15a) and (15b).

(15) a.



b.



Consider first (15a). Since the article consists of a single word, the CRD of the noun phrase begins with the word *the*. If we assume that the relative pronoun *who* serves as a MNCC for the relative clause, then the relative pronoun will be the MNCC for the last daughter of the noun phrase, so that the CRD will end with the relative pronoun. Hence, the CRD for (15a) will be *the man who*. On the other hand, the CRD for (15b) will consist of the entire noun phrase *the man who told me that Smith left*, since the first and last daughters consist of single words. Hence Hawkins' principle correctly predicts a preference for the word order in (15a) over that in (15b).

There is one problem, however, with this explanation in terms of Hawkins' Principle. Namely, it implies a preference for Art-N-Rel order over N-Rel-Art order regardless of whether the language is VO or OV, since the argument makes no reference to the order of object and verb. But what we are trying to explain is why ArtN order is more common in VO languages while NArt order is more common in OV languages. The point is a significant one since the two orders of relative clause and noun are about equally common among OV languages (Dryer 1992a, 2005b). The appeal to Hawkins' Principle would seem to predict that that OV languages which are NRel should show the same preference for ArtN order as that found in VO languages. However, it turns out that this is not the case. To the contrary, OV&NRel languages exhibit the same (weak) preference for NArt order as that found in OV&RelN languages. An example from an OV&NRel language that employs the N-Rel-Art word order of (12) is shown in (16).

(16) Lepcha (Tibeto-Burman; Mainwaring 1876: 43)

máro	[to	nun	zuk]	re
person	who	ERG	do	DEF

'the person who did it'

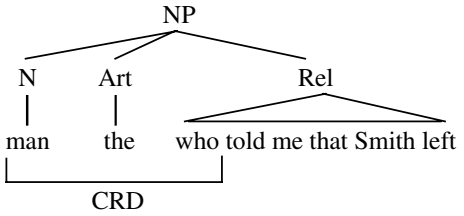
Hawkins' Principle does correctly predict NArt order in a language that is OV&RelN, since placing the article at the beginning of the noun phrase would mean that the CRD for the noun phrase would start at the beginning of the noun phrase, so that the CRD would be the entire noun phrase, analogous to (12) (but with the mirror image word order), while placing the article at the end of the noun phrase would mean that the CRD for the noun phrase would not start until the noun was encountered, after the relative clause.

One possible explanation for the high occurrence of N-Rel-Art order in OV languages is that the article helps processing in that it signals the end of the relative clause. This is more useful in OV languages than in VO languages because an object noun phrase containing a relative clause will end inside the main clause in an OV language but at the end of the main clause in a VO language. Signaling the end of a subordinate clause within a main clause is more useful for processing than signaling the end of a subordinate clause that is also the end of a sentence.[2] The above discussion considers two possible positions for an article in an NRel language, one at the beginning of the noun phrase, with Art-N-Rel order, the other at the end of

[2] This idea was suggested to me by Lea Brown.

the noun phrase, with N-Rel-Art order. But if we assume flat structures, then there is a third logical possibility for the placement of the article; namely it might occur immediately after the noun, as in (17).

(17)



This word order is found in Koyraboro Senni, as in (18), where there are actually two morphemes that could be treated as definite articles, the definite clitic *=oo* and the anaphoric demonstrative *din*.

(18) hondu beer=oo din [kaŋ n=ga diy-aa]
dune big=DEF.SG ANAPH REL 2SG.SUBJ=IMPF see-3SG.OBJ
'the great dune that you see' (Heath 1999: 244)

But this order is less common than NRelArt order. In (19) are listed languages of the two types NRelArt and NArtRel, first for VO languages and then for OV languages. This shows NRelArt outnumbering NArtRel by 13 to 6 among VO languages and by 11 to 3 among OV languages.

(19) VO&NRelArt: 13 languages: Busa, Mupun, Maybrat, Kera, Gimira, Sobei, Sudest, Bali-Vitu, Sisiqa, Longgu, Fongbe, Bagirmi, Sundanese

VO&NArtRel: 6 languages: Koyra Chiini, Tetun, Kaulong, Linda, Nadrogā, Paamese

OV&NRelArt: 11 languages: Tshangla, Kanuri, Tubu, Maba, Seri, Busa, Arrernte (Mparntwe), Runga (Maban), Lepcha, Takia, Dogon

OV&NArtRel: 3 languages: Koyraboro Senni, Kairiru, Ute

We must look for some factor other than Hawkins' Principle to predict that the NArtRel word order in (17) tends to be avoided, since the length of the CRD is the same as ArtNRel in (15a) and shorter than the length of the CRD with NRelArt order in (15b).

The relative infrequency of structures like that in (17) can be seen as reflecting a generalization that is independent of word order correlations. Namely, crosslinguistically, articles tend to occur on the periphery of noun phrases, either as the first word, or as the last word. Hence, if we can explain why this is the case, we would have an explanation for why VO languages prefer the word order in (15a) over those in both (15b) and (17), and hence why ArtN order is preferred to NArt order in VO languages.

Now some linguists have an easy answer to why articles tend to occur on the periphery of noun phrases. Namely, those linguists who believe in more hierarchical structures than those I have been assuming claim that articles combine with N. If

articles combine with N$^{|}$s (or combine with NPs to form DPs), then structures like (17) are impossible, or at the very least, they involve a discontinuous N, and would tend to be avoided for that reason. However, since the goal of this paper is to discuss how to explain the word order correlations if we assume flat structures, that line of explanation is not open to me.

I would suggest as an alternative that the position of articles on the periphery of noun phrases is simply one manifestation of the more general principles governing the order of noun modifiers, that leads to adjectives tending to be closer to nouns than numerals, demonstratives, and articles, explaining why (20b) and (20c) are ungrammatical in English and to certain preferred orders among different sorts of adjectives, explaining why (20a) is preferred over (20d) .

(20) a. the three large black dogs
b. *three the large black dogs
c. *three large black the dogs
d. ??the three black large dogs

While I know of no completely satisfactory account of these preferences (but see Bache 1978, Posner 1986, Sproat and Shih 1988), they seem to involve some principle according to which words that denote more inherent properties of the referent of the noun phrase tend to occur closer to the noun. Thus in (20a), the adjectives denote properties that the dogs have had for a long time, while the numeral *three* denotes a property that has been true only as long as the dogs are together, while the pragmatics of the definite article do not denote a property of the dogs at all, but simply their status in the discourse. I suspect it is some principle like this that explains why articles tend to occur on the periphery of noun phrases.

I should note that even if one accepts the idea that some principle like this governs the order of noun modifiers, one might claim that this principle simply reflects a deeper fact that the semantic structure of noun phrases with multiple modifiers is such that in a noun phrase like (20a) with an article, a numeral, and an adjective, the adjective plus noun forms a semantic unit to the exclusion of the article and numeral and that the numeral plus adjective plus noun forms a semantic unit to the exclusion of the article. According to this approach, the noun phrase *the three large dogs* has the semantic structure *[the [three [large dogs]]]*. The fact that languages avoid the word order in *the large three dogs* would be explained on this approach by saying that *three dogs* does not form a semantic unit to the exclusion of *large*. A proponent of this idea could argue that this explains why modifiers denoting inherent properties tend to occur closer to the noun, rather than the other way round.

While this approach may have some merit, it denies the premise of this paper, that the semantic structure of noun phrases is flat. In other words, I am trying to explain the word order correlations if we assume flat structures. If this premise is false, if articles do in fact combine with semantic units, then the BDT will account for the fact that articles are verb patterners. In other words, the premise may be false, but then the fact that the order of article and noun correlates with the order of verb and object would be correctly predicted by the BDT.

But there is another reason why it is not crucial whether the semantic structure of noun phrases is more hierarchical than I am assuming in this paper. Namely, what I

am considering a semantic unit is a set of words which the hearer must recognize as a semantic unit in order to understand the noun phrase. But even if noun phrases are viewed as having an hierarchical semantic structure, it is not clear that the hearer must recognize the various units in this hierarchical structure as semantic units in order to understand the noun phrase. In other words, even if a noun phrase is viewed as having a nested hierarchical semantic structure, the semantic units in this hierarchical structure are not fundamental semantic units in the way that noun phrases and clauses are. The claim of this paper is that it is only these fundamental semantic units that must be recognized as semantic units if a sentence is to be understood.

Before leaving discussion of articles, there is one additional correlation that is worth discussing. Namely, not only do VO and OV languages differ in the order of article and noun, but they also differ in how frequently they employ articles. The data in (21) gives the number of languages in my database that have articles and the number that do not have articles for both OV and VO languages.

(21)

	OV	VO
Has articles	183	312
Does not have articles	101	57

What (21) shows is that languages with articles outnumber those without by less than 2 to 1 among OV languages, but by over 5 to 1 among VO languages, so that VO languages are more likely to have articles than OV languages.[3]

Interestingly, Hawkins' theory provides a possible explanation for this difference in the frequency of articles. Namely, as discussed above, both nouns and articles serve as MNCCs for noun phrases. Now an important difference between parsing right-branching languages and parsing left-branching languages that is predicted by Hawkins' theory is that in parsing right-branching languages, the CRD of a constituent will tend to be shorter the more quickly the last daughter of that constituent is recognized. It is thus an advantage in a right-branching language to have more than one MNCC for a constituent, since the more MNCCs there are, the more quickly the constituent will be recognized, and the shorter the CRD of the mother of that constituent will be. Conversely, in a left branching language, Hawkins' theory predicts that the CRD of a constituent will be shorter, the later the first daughter of that con-

[3] There is one feature of the numbers in (21) that is highly misleading and purely an artifact of the way in which the data was collected. Namely, the numbers in (21) suggest that the majority of languages of the world have articles. But this is probably not the case. Rather the higher numbers for languages with articles in (21) simply reflects the fact that it is easier to infer from a grammatical description of a language that it has articles than it is to infer that it does not have articles (since the article might be optional). Namely, if I found evidence in a description that a language has articles, then I coded it as having articles. But a description had to be fairly detailed and fairly clearly written for me to conclude that the language lacks articles. If a description was brief, or if it was less clearly written, or if for some reason I had only limited time to examine it, I would not code the language as lacking articles in the absence of finding evidence of articles. It is only for this reason that the numbers for languages with articles in (21) are greater than those without. It is my educated guess that languages with articles amount to at most half the languages of the world, and probably somewhat less than half.

stituent is constructed. In other words, while someone processing a right-branching language wants to recognize constituents as quickly as possible, someone processing a left-branching language wants to recognize constituents as late as possible. But having multiple MNCCs is thus a disadvantage in a left-branching language, because having multiple MNCCs means that the constituent with multiple MNCCs will be recognized earlier. But since a language with articles will have two MNCCs for noun phrases while a language without articles will have only one, this means that Hawkins' theory predicts that right-branching languages are more likely to have articles than left-branching languages, and since VO languages are generally right-branching while OV languages are generally left-branching, this predicts that VO languages are more likely to have articles than OV languages.

5 The Position of Complementizers

The second pair of elements that I will discuss is the order of complementizer and clause. Some languages have clause-initial complementizers, like English, as in (22), while other languages have clause-final complementizers, like Japanese, as in (23).

(22) John knows [***that*** we have left].

(23) OV&FinalComp: Japanese

John	wa	[nihongo	ga	muzukasii	***to***]	it-ta
	TOPIC	[Japanese	SUBJ	difficult	***COMP***]	say-PAST

'John says [that Japanese is difficult].'

Since I did not present data on the order of complementizer and clause in Dryer (1992a), I present in (24) data for this, using the format in that paper, where the numbers denote numbers of genera rather than numbers of languages, although I give the total number of languages of each sort in the rightmost column.

(24)

	Africa	Eurasia	SEAsia&Oc	Aus-NewGui	NAmer	SAmer	Total	#Lgs
OV&FinalComp	2	[5]	[3]	1	[2]	[1]	14	27
OV&InitComp	[6]	4	1	[3]	0	0	14	22
VO&FinalComp	0	0	0	0	0	0	0	0
VO&InitComp	[23]	[9]	[13]	[4]	[10]	[4]	63	140

What (24) shows is that the two positions of complementizers are about equally common in OV languages, but that all of the VO languages for which I have data on the position of complementizers place the complementizer at the beginning of the clause. The data in (24) shows 63 genera containing languages with clause-initial complementizers; the total number of languages is 140.

Note that although the two positions of complementizers are about equally common in OV languages, it is still the case that the order of complementizer and clause correlates with the order of object and verb and that complementizers are verb patterners since OV languages use clause-final complementizers significantly more

often than VO languages. The asymmetry in (24) requires some further explanation, but it is an instance of a broader generalization, and that is that while structures of the form phrase+word (e.g., NP + Postposition) are as common as structures of the form word+phrase (e.g. Preposition + NP), structures of the form clause+word are much less common crosslinguistically than structures of the form word+clause (Dryer 1992b). Another instance of this generalization is that prenominal relative clauses are much less common crosslinguistically than postnominal relative clauses.

The example in (23) above from Japanese illustrates an OV languages where the complementizer is clause-final; the examples in (25) from Supyire illustrate an example of an OV language in which the complementizer is clause-initial; (25a) illustrates the OV word order, while (25b) illustrates the clause-initial complementizer.

(25) OV&InitComp: Supyire (Gur, Niger-Congo; Mali; Carlson 1994: 339, 423)

a. u náá naŋiyááyi kàni ɲya
3SG REM.PAST wild.animals.DEF only see
'he saw only the wild animals'

b. mìi a lì ɲyɛ̀ [na u a kàrè]
1SG PERF it see [COMP 3SG PERF go]
'I saw [that he had gone]'

But there is another important difference between Japanese and Supyire. Not only do these two languages differ in the position of complementizers, but they also differ in the position of the complement clause: in Japanese, the complement clause precedes the main verb, while in Supyire the complement clause follows the main verb. And this difference in the position of the complement clause among OV languages correlates very strongly with the position of complementizers, in that we have the two generalizations stated in (26).

(26) a. OV languages in which the complement clause precedes the verb normally have clause-final complementizers rather then clause-initial complementizers.
b. OV languages in which the complement clause follows the verb normally have clause-initial complementizers rather then clause-final complementizers.

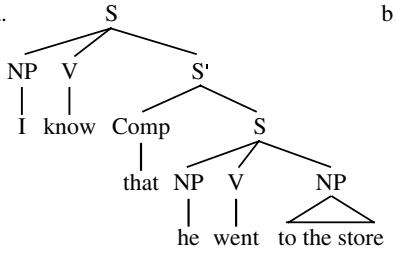
The specific data I have on this is given in (27).

(27) *OV, preverbal complement clause, final complementizer:* 12 languages:
Dogon, Orkhon Turkic, Japanese, Ainu, Kannada, Hayu, Mao Naga, Angami, Lai Chin, Bawm, Amele, Slave
OV, preverbal complement clause, initial complementizer: 1 languages:
Harar Oromo
OV, postverbal complement clause, final complementizer: 1 language:
Khoekhoe (aka Nama)
OV, postverbal complement clause, initial complementizer: 18 languages:
Mauka, Supyire, Tunen, Latin, Hindi, Punjabi, Marathi, Wakhi, Pashto, Persian, Tajik, Wakhi, Turkish, Tsova-Tush, Gapapaiwa, Tawala, Sare, Djapu

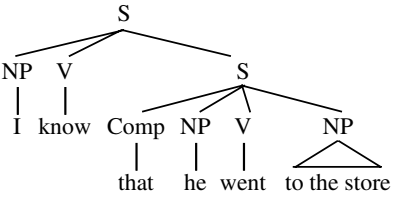
Only two languages in (27), Harar Oromo and Khoekhoe, do not conform to (26), while the other 29 languages do conform. I will return to these generalizations shortly.

Let us turn to the issue of explaining why complementizers are verb patterners, occurring at the beginning of the clause significantly more often in VO languages than in OV languages. In Dryer (1992a), I explained this in terms of the BDT, assuming a structure like that in (28a) in which the complementizer combines with an S to form an S'. On the other hand, if we assume a flatter structure like that in (28b) (which was the usual constituent structure assumed in generative grammar before around 1970), then the complementizer is no longer combining with a phrasal constituent.

(28) a.



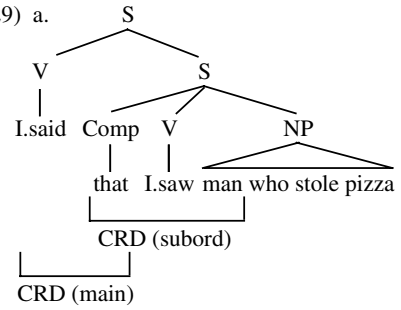
b.



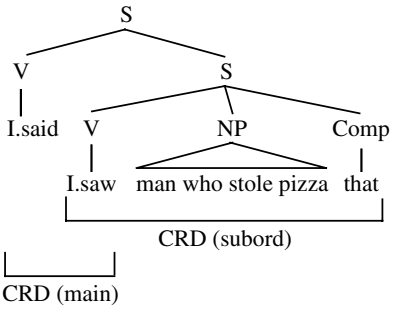
As a result, the BDT fails to predict that a VO language will place the complementizer at the beginning of the clause, rather than at the end.

Hawkins' Principle, however, does account for this. Compare the two structures in (29), where (29a) illustrates a clause-initial complementizer and (29b) a clause-final complementizer. The structures given represent the subject as being expressed in the verbal morphology, rather than by a separate pronoun, since this is the way the majority of languages express pronominal subjects, and this proves useful in comparing a number of different possible structures below, all intended as ways of expressing the meaning 'I said that I saw the man who stole the pizza'. I include the CRDs for both the complement clause and for the main clause, since considering both of these will be relevant below when I consider the possibilities in OV languages.

(29) a.



b.

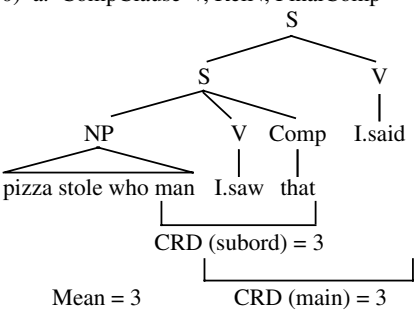


Hawkins' Principle does account for the preference for (29a) over (29b): although the CRDs for the main clause are the same in these two structures (both consisting

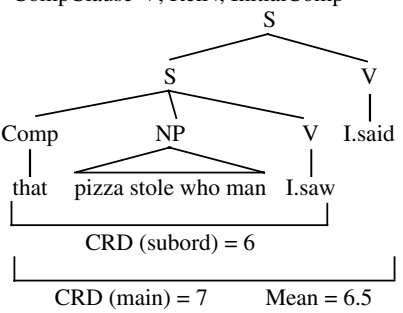
of only two words), the CRD for the complement clause is shorter in (29a) than it is in (29b). The CRD for the subordinate clause in (29b) is the entire clause, while in (29a) it only includes the first word of the last daughter of the clause.

Now consider the situation in an OV language. There are in fact eight types of OV languages to consider, based on three binary variables, position of complementizer in clause, order of complement clause with respect to the verb, and the order of relative clause and noun. For all three of these variables, both orders are well attested among OV languages. These eight possibilities are given in (30) to (33). The CRDs for the subordinate clause and the matrix clause, as well as the length in words of each CRD and the mean of these two values is given for each tree.

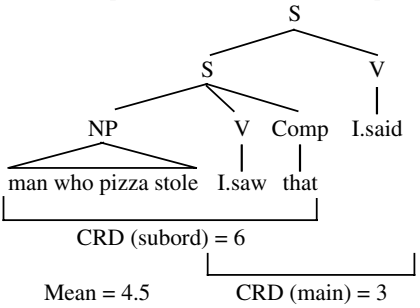
(30) a. CompClause-V, RelN, FinalComp



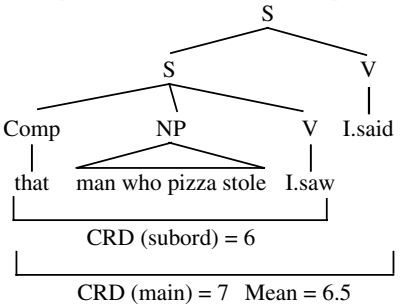
b. CompClause-V, RelN, InitialComp



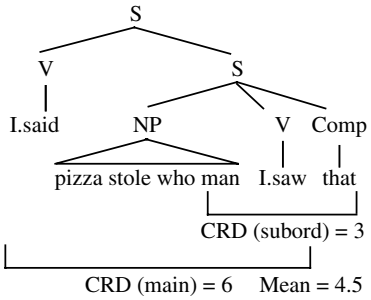
(31) a. CompClause-V, NRel, FinalComp



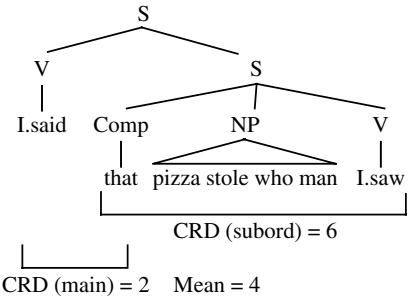
b. CompClause-V, NRel, InitialComp



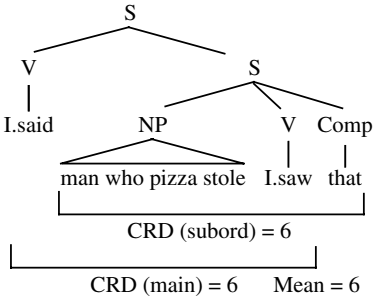
(32) a. V-CompClause, RelN, FinalComp



b. V-CompClause, RelN, InitialComp



(33) a. V-CompClause, NRel, FinalComp



b. CompClause-V, NRel, InitialComp

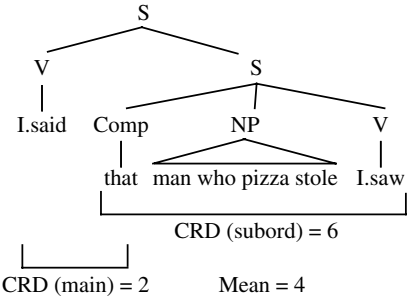


Table 4 summarizes the mean length of the CRDs for the eight structures in (30) to (33), and lists attested languages of each sort.[4]

Table 4 shows that although the order of relative clause and noun has some effect, the basic generalization is that final complementizers lead to shorter CRDs if the complement clause precedes the verb while initial complementizers lead to shorter CRDs if the complement clause follows the verb. Hence Hawkins' Principle cor-

Table 4 Position of complement clauses, complementizers, and relative clauses in OV languages

		FinalComp		Initial Comp	
		Mean length of CRD	Attested languages	Mean length of CRD	Attested languages
ClauseV	RelN	3	8 languages: Japanese, Ainu, Orkhon Turkic, Kannada, Hayu, Mao Naga, Lai Chin, Bawm	6.5	Zero languages
	NRel	4.5	1 language: Amele	6.5	1 language: Harar Oromo
VClause	RelN	4.5	1 language: Khoekhoe (aka Nama)	4	3 languages: Marathi, Turkish, Tsova-Tush, Sare
	NRel	6	Zero languages	4	6 languages: Pashto, Persian, Tajik, Wakhi, Gapapaiwa, Tawala

[4] Some of the languages listed in (27) above are not included in Table 4 because they employ either internally-headed or correlative relative clauses.

rectly predicts that VO languages should have initial complementizers, since complement clauses in VO languages invariably follow the verb (Dryer 1980) and that OV languages will have final complementizers if the complement clause precedes the verb and initial complementizers if the complement clause follows the verb. In addition, the numbers of attested languages of each of the types in Table 4 matches the predictions of Hawkins' Principle: the three types whose mean CRD is 4 or less are exactly the types for which more than one language is attested.

The numbers in (24) above show another difference between OV and VO languages other than the position of complementizers. Namely, (24) includes data for 49 OV languages with complementizers but for 140 VO languages with complementizers. This difference reflects the fact that VO languages employ complementizers more often than OV languages do. This is analogous to the fact discussed in Section 4 that VO languages have articles more often than OV languages do. And the explanation for the lower frequency of articles in OV languages in terms of Hawkins' Principle also applies to the lower frequency of complementizers in OV languages. Namely, complementizers, like verbs, are MNCCs for clauses. Having multiple MNCCs for a given category is an advantage in right-branching languages but a disadvantage in left-branching languages. This is reflected in the trees in (30) to (33) in that the shortest CRDs for preverbal complement clauses contained three words, while the shortest CRDs for postverbal complement clauses contain only two words. This is because with preverbal complement clauses, both the complementizer and the verb in the complement clause are part of the CRD for the matrix clause, since as soon as one of them is encountered, the complement clause can be constructed by the parser. With postverbal complement clauses, however, once the complementizer is encountered, the complement clause is constructed, so the verb in the complement clause will not be part of the CRD for the matrix clause.

As with articles, Hawkins' Principle only partially explains why complementizers are verb patterners. If we assume flat structures in which the complementizer is simply one of the constituents of the clause, what we need to explain is why complementizers do not occur inside clauses. If one considers what (29a) would look like with the complementizer immediately before the verb or immediately after the verb, one can see that the length of the CRD would be the same as in (29a). But languages with clause-internal complementizers are rare. As was the case with articles, complementizers normally occur at the periphery of clauses, either at the beginning or at the end. To fully explain why complementizers are verb patterners, we would need to explain this generalization.

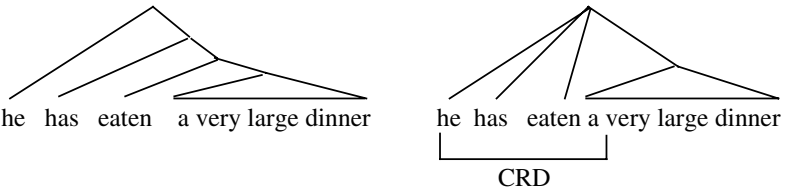
Once again, one obvious way to explain why complementizers occur at clause boundaries would be to say that they combine semantically and/or syntactically with clauses, as in (28a) above, where they combine with an S to form an S′. However, can we explain it without making such a claim? One possibility is that one of the functions of complementizers is to signal clause boundaries. Since identifying which words go together in the same clause is essential for understanding a sentence, having an overt signal of a clause boundary is advantageous for sentence processing. A complementizer inside a clause would clearly not be helpful in signaling a clause boundary.

The hypothesis that complementizers signal clause boundaries also provides an additional explanation for why languages would prefer (29a) over (29b): placing a complementizer at the end of a sentence is not going to be helpful in signaling a clause boundary within a sentence. This provides an additional explanation for why languages with preverbal complement clauses employ clause-final complementizers and why languages with postverbal complement clauses employ clause-initial complementizers.

6 The Order of Auxiliary Verb and Main Verb

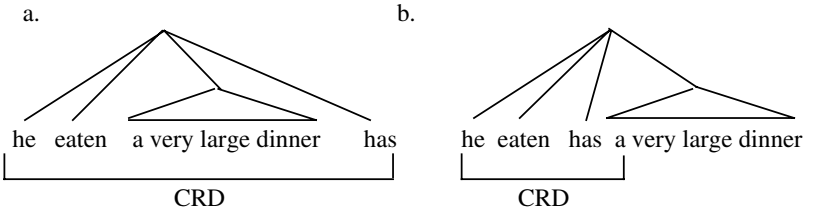
My discussion of this pair of elements will be fairly brief. The primary conclusion is that there is no obvious explanation for why auxiliary verbs tend to precede the main verb in VO languages but follow in OV languages. The BDT explains the correlation only if we assume a structure like that in (34a), where the auxiliary verb is combining with a verb phrase *eaten a very large dinner*. If we assume the flatter structure in (34b), the auxiliary verb is no longer combining with a phrasal category.

(34) a. b.



Hawkins' Principle does correctly predict that VO languages will prefer the structure in (34b) to that in (35a), with the auxiliary verb at the end of the clause, since in (35a), the CRD for the clause is longer, since it includes the entire clause. However, it does not explain why the word order in (35b) is avoided, since the length of the CRD in (35b) is the same as in (34b).

(35) a. b.



Nor do the sorts of factors that cause articles and complementizers to occur at the periphery of their mother constituents seem to be relevant here, since if we assume the flat structures in (34b) and (35b), the auxiliary verb is not in peripheral position in either structure, nor is it the case that auxiliary verbs tend to occur at the periph-

ery of clauses, since in SVO languages they typically occur after the subject and immediately before the main verb.

My only suggestion is that explanations in terms of grammaticalization are especially plausible with auxiliary verbs since the processes of grammaticalization of main verbs to auxiliary verbs is one of the most frequent forms of grammaticalization. Under this approach, their position relative to the main verb reflects their position as main verb prior to the grammaticalization. Note that if this is the correct explanation, then the word order correlations are not a unified phenomenon, some being due to processing factors, some due to grammaticalization, and perhaps some due to other factors, and perhaps some due to a combination of factors. While a single explanation might seem a priori preferable, it is certainly undesirable if it is incorrect.

7 Conclusion

For reasons of space, I have not been able to discuss the other problematic pairs listed in Table 3 above, but let me make some very brief suggestions. If adpositions are simply constituents of noun phrases, then the problem of explaining their correlation is rather similar to the issues surrounding the order of article and noun, discussed in Section 4, since they, like articles, would serve as MNCCs for noun phrases. Grammaticalization also appears to be a factor in explaining the position of adpositions in many languages, so this is a case where two different sorts of factors may be working together. The situation with plural words may also be similar to that with articles, but this is less obvious, since unlike articles and adpositions, they frequently do not occur at the periphery of noun phrases. The explanation for the position of negative auxiliaries is presumably the same as that with tense-aspect auxiliaries. The situation with adverbial subordinators is similar to that with complementizers in many respects, and they plausibly serve as signals of clause boundaries (in addition to signalling a specific semantic relationship between the subordinate clause and the main clause). As for question particles, if we can explain why they tend to occur at the beginning or end of the sentence (perhaps in terms of the fact that they are modifying the sentence as a whole), then processing considerations would predict that in a verb-final language, they would tend to occur at the end of the sentence, since that would result in a shorter CRD for the sentence than if they occurred at the beginning of the sentence.

In conclusion, it appears that for some pairs of elements whose status as correlation pairs is not explained by the BDT under assumptions of flat constituent structure, the correlations can be explained, at least partly, by a combination of Hawkins' Principle with factors that lead certain sorts of words to occur on the periphery of their mother constituents. But there are other pairs, like auxiliary verb and main verb, that do not seem explainable in this way. Further examination of the other correlation pairs that become problematic for the BDT under assumptions of flat constituent structure is still needed.

References

Bache, Carl. 1978. *The order of premodifying adjectives in present-day English*. Odense: Odense University Press.

Carlson, Robert. 1994. *A grammar of Supyire*. Berlin: Mouton de Gruyter.

Culicover, Peter W. and Ray Jackendoff. 2005. *Simpler syntax*. Oxford University Press.

Dixon, Robert M. W. 1997. *The rise and fall of languages*. Cambridge: Cambridge University Press.

Dryer, Matthew S. 1980. The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25: 123–195.

Dryer, Matthew S. 1992a. The Greenbergian word order correlations. *Language* 68: 81–138.

Dryer, Matthew S. 1992b. Adverbial subordinators and word order asymmetries. In *Performance principles of word order*, eds. John A. Hawkins and Anna Siewierska, 50–67. EUROTYP Working Papers, European Science Foundation.

Dryer, Matthew S. 2005a. Relationship between order of object and verb and order of adjective and noun. In *World atlas of language structures*, eds. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 394–397. Oxford University Press.

Dryer, Matthew S. 2005b. Relationship between order of object and verb and order of relative clause and noun. In *World atlas of language structures*, eds. Martin Haspelmath, Matthew S. Dryer, David Gil, and Bernard Comrie, 390–393. Oxford University Press.

Frazier, Lyn. 1978. On comprehending sentences: syntactic parsing strategies. University of Massachusetts Ph.D. dissertation.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed. Joseph, Greenberg, 73–113. Cambridge, Mass: MIT Press.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Heath, Jeffrey. 1999. *A grammar of Koyraboro (Koroboro) Senni*. Köln: Rüdiger Köppe Verlag.

Posner, Roland. 1986. Iconicity in syntax: the natural order of attributes. In *Iconicity: Essays on the nature of culture: festschrift for Thomas A. Sebeok on his 65th birthday*, ed. P. Bouissac, M. Herzfeld and R. Posner, 305–337. Tübingen: Stauffenberg.

Sproat, Richard and Chillin Shih. 1988. Prenominal adjectival ordering in English and Mandarin. In *Proceedings of the North Eastern Linguistic Society* (NELS) 18: 465–488.

Van Valin, Robert, and Randy LaPolla. 1997. Syntax: Structure, meaning, and function. Cambridge: Cambridge University Press.

Universals and Semantics

Denis Delfitto*

Abstract This contribution proposes a view of linguistic semantics as a set of mental computations defined on a suitably restricted inventory of interpreted features borrowed from conceptual structures external to the language organ. These features enter both a lexical and a syntactic computation. Semantic universals can be identified regarding the nature of these featural primitives, the nature of the lexical computation (involving both formal and substantive universals), the nature of the mapping between syntactic categories and notional categories, the role of grammatical features in pre-encoding interpretive operations. It is argued that consensus about semantic universals can be reached cutting across the artificial divide between functionalist and formalist approaches to human language.

Keywords Semantic universals · lexical and syntactic computations · set-theoretic properties · mapping conditions · grammatical features · morphosyntactic pre-encoding

1 Introduction

In the debate about universals, it is common to distinguish between a typological approach, according to which universals are identified with constraints on cross-linguistic variation defined in general cognitive terms and directly bearing on the communicative function of language, and a formalist approach, according to which universals are the defining property of language as a mental organ and represent the logical solution to the 'poverty of the stimulus' problem (for a recent overview, see Mairal & Gil 2006). In a formalist perspective, knowledge of meaning offers quite convincing examples of poverty of the stimulus. Consider for instance the paradigm

D. Delfitto
University of Verona

* I am indebted to Chiara Melloni and two anonymous reviewers for their written comments on a previous version of this contribution.

S. Scalise et al. (eds.), *Universals of Language Today*, 209–224,
DOI: 10.1007/978-1-4020-8825-4_11,

in (1) (Fillmore 1982, Jackendoff 2002), with (1d) traditionally assigned a more degraded status than the other variants in (1):

(1) a. Bill climbed (up) the mountain
b. Bill climbed down the mountain
c. The snake climbed (up) the tree
d. ?*The snake climbed down the tree

Since the analysis of verb meanings in English reveals a clustering of motion and path on one side and motion and manner on the other side (with complementarity between path and manner; cf. Levin and Rappaport 1992), it is arduous to explain how an arbitrary set of primary linguistic data can trigger the knowledge of the sort of disjunctive lexical meaning of 'climb' that seems required in order to account for the complex pattern of judgments in (1) ('climb' must involve either a clambering manner of motion *or* an upwards direction). This kind of evidence can be used as a strong argument in favor of the view that there must be concealed universal conditions governing the construction of the primitive building blocks that correspond to the lexical entries of a given language. On a similar vein, one might argue that there is evidence for universal conditions on the application of lexical operations. Consider in this respect the so-called causative/intransitive alternation exemplified in (2):

(2) a. Bill broke the window
b. The window broke
c. Bill killed his wife
d. *His wife killed

The kind of valency-reduction operation attested by the transition from causative (2a) to intransitive (2b), intuitively consisting in the suppression of the agent theta-role, cannot successfully apply to (2c), yielding (2d), roughly meaning that Bill's wife died as a result of Bill's killing. It is tempting to explain this contrast in terms of a universal ban on the suppression of the agent theta-role whenever the latter is conceptualized as [+mental +cause] (as is the case with 'kill') and not simply as [+cause], as is the case with 'break' (where the 'agent' is not necessarily animate, as in 'the storm broke the window'; cf. especially Reinhart 2002). There is compelling psycholinguistic evidence that speakers, when using causative verbs, single out events involving direct and intentional causation (a person *waved a flag* only when she shook a flagpole, not when she raised the flag on a windy day; cf. Wolff 2003 and the discussion in Pinker 2007). In English and many other languages, verbs referring to human actions do not take part in the causative alternation (**Bill laughed Mary*, **Bill cried Mary*), on a par with verbs referring to physical events in which an object emits something, such as a light, a sound, a substance (**glow a light*; see Levin 1985, Pinker 2007), strongly suggesting that the speaker conceptualizes these events as involving internal, inherent causation. Of course, the claim here is not that there are no exceptions to this generalization. However, to put it in Pinker's words, "[the causative and other related alternations, such as the locative alternation] repeatedly turn up in unrelated languages and language families all over the world,

suggesting that people's language-forming abilities, faced with the need to communicate certain kinds of ideas, are channeled to rediscover these constructions". More particularly, "languages with a documented causative alternation number in the hundreds, and many surveys have ferreted out their common properties" (Pinker 2007: 79, and the references cited therein).

Knowledge of meaning provides thus a strong motivation for underlying constraints on linguistic knowledge, both in the form of substantive universals (such as the inventory of primitive semantic ingredients for constitution and structuring of lexical meaning) and in the form of abstract universals (such as the formal conditions governing lexical operations of valency-reduction). It is an empirical matter to decide whether the proposed universals reflect the application of general cognitive constraints (phylogenetically related to functional properties of language, as in Newmeyer 2005) or of a well-defined set of language-specific conditions, as in the formalist models of syntactic knowledge. Investigation into semantic universals is in fact made more challenging, in my opinion, by the absence of an a *priori* bias towards either language-specific or non-language-specific principles, as is to be expected, under reasonable assumptions, for a domain of inquiry aiming to study the interaction between the computational system underlying language (Chomsky 1995) and the interpretive systems at its boundaries, crucially including conceptual structure.

2 Semantics as a Linguistic Module

Under the theory of meaning adopted here, there is an interesting correspondence between syntactic and semantic computations, that should be emphasized. If the inventory of formal features that enter the syntactic computation is a subset of the full inventory of features that constitute lexical entries, it is reasonable to assume that the inventory of interpreted features that feed the semantic computation represents a subset of the full inventory of notional features that constitute conceptual structure and, more generally, the systems of interpretation. In this sense, semantics is a kind of broad syntax, manipulating a restricted set of interpreted features corresponding to selected parts of conceptual structure (Chomsky 2000), provided one is disposed to concede that it is broad syntax so defined, and not *narrow* syntax in Chomsky's sense, to constitute the core of the language faculty (Jackendoff 2002). An example of semantic computation is the lexical operations giving rise to the causative/unaccusative alternations shown in (2a–b) and, more generally, to the application of causativization or decausativization strategies starting from a single lexical entry with well-defined thematic properties (more on this in Section 3 below). Another case in point is the manipulation of formal objects endowed with set-theoretic properties, as when the set-theoretic properties assigned to determiners appear to be relevant for the licensing of negative polarity items (cf. Ladusaw 1979 and much subsequent literature). Consider for instance that "any" is correctly licensed by "every" in (3a) but not in (3b):

(3) a. Every student who praised **any** professor succeeded
b. *Every student who succeeded praised **any** professor

Under the hypothesis that "any" needs to be in the scope of a monotone decreasing operator and that the determiner "every" expresses a relation between two sets A and B (whereby A is included in B), the grammaticality contrast between (3a) and (3b) is directly explained by the fact that "every" is monotone decreasing in the first set (i.e. the set A, corresponding to the nominal restriction of the determiner) but not in the second (i.e. the set B, corresponding to the VP). This is proved by the observation that there are valid entailments going from A to subsets of A, but there is no valid entailment going from B to subsets of B, as shown in (4):

(4) a. "every student who praised the professor succeeded" *entails*
"every student who verbally praised the professor succeeded"
b. "every student who succeeded praised the professor" *does not entail*
"every student who succeeded verbally praised the professor"

Since the polarity item "any" is contained within the linguistic expression corresponding to A in (3a), whilst it is contained in the linguistic expression corresponding to B in (3b), the condition stating that "any" must find itself in the scope of a monotone decreasing operator is satisfied in (3a) but not in (3b), accounting for the degraded status of "any" in (3b).

Significantly, the relevance of set-theoretic computations for the satisfaction of grammaticality conditions arguably extends beyond nominal syntax and the semantics of determiners. It is monotonicity that accounts in fact for the unexpected grammaticality contrast detected in (5) (see Higginbotham 1988 for a full discussion):

(5) a. John left before anyone else
b. *John left after anyone else

The case in (5) is particularly telling: It is only if we are ready to concede that the semantic computation applies to objects endowed with set-theoretic properties that superficially similar lexical items (*after* and *before* in the case under scrutiny) disclose deep concealed differences (see also Delfitto 2003).

I propose thus that the semantic computation applies to abstract objects built up from relevant fragments of conceptual structure (and, more generally, the Conceptual-Intentional systems), whose formal properties (possibly expressed in a set-theoretic vocabulary) may bear on the notion of logical consequence (for further discussion, the reader is referred to Section 5). A noticeable consequence of the view adopted here is that language must be fully distinguishable from conceptual structure and from the interpretive systems into which it is mapped. However, there is clear evidence according to which conceptual structure feeds language. For instance, it is the speaker's ability to structure verb meaning, within a 'language of thought', in terms of subtle semantic properties such as manner of motion, change of state or mental state that is responsible for verb behavior in causative (and locative) alternation phenomena (i.e. for the mapping into syntactic structures), in complete disregard of other interpretive ingredients that may well appear in dictionary

definitions of verb meanings (cf. the discussion in Pinker 2007, Chapter 2). On these grounds, we propose that even from a semantic perspective, language presupposes thought (as we saw above, the featural system that feeds semantic computations is certainly related to the C-I system), whereas thought does not presuppose language (one may easily conceive of entities endowed with a C-I system but completely deprived of the sort of semantic computations briefly sketched above, applying to lexical items and/or syntactic structures and delivering valency-reduction, licensing of polarity items, etc.).

A lexicon involves thus a selection of proper subparts of conceptual structure (see also Lieber 2004's distinction between 'skeleton' and 'body') and gives rise, under the view adopted here, to two distinct forms of linguistic computation: a lexical computation and a syntactic computation. It follows that there are two distinct forms of semantic computation: one is pre-syntactic (and coincides with the lexical computation) and one is post-syntactic (coinciding with phrasal semantics). It is to the semantic universals that can be identified in each of these two distinct domains that we now devote our attention.

3 Lexical Semantics: Formal and Substantive Universals

Starting with formal universals, there is widespread consensus that there is a unique system of lexical representations underlying some of the morphological distinctions detected within the verbal system. Consider in particular the hypothesis according to which distinct thematic forms are universally related to a single thematic entry by means of a well-defined set of lexical operations, informally stated in (6):

(6) **Universal** *(Lexicon Uniformity Hypothesis)*:
Each verb-concept corresponds to one lexical entry with one thematic structure. The various thematic forms of a given verb are derived by lexical operations from one thematic structure

On these conceptual grounds, one can argue that the set of *Aktionsart* distinctions originally identified by Vendler 1967 for English proves actually to have a crosslinguistic validity, accounting for the pervasiveness of the two symmetrical morphological strategies informally exemplified in (7), relating states, accomplishments and causatives (Van Valin 2006: 173):

(7) a. *Causativization* (Tibeto-Burman languages, Quechua, etc.)
State → Accomplishment/Achievement → Causative
ba (big) → *tə-ba* (become big) → *tə-ba-z* (cause to become big) (*Qiang*)

b. *Decausativization* (Russian, French, etc.)
Causative → Accomplishment/achievement → State
razbit' (break) → *razbit'sja* (break, intr.) → *razbityj* (broken) (*Russian*)
briser (break) → *se briser* (break, intr.) → *brisé* (broken) (*French*)

(7a) illustrates the case where the base form of the verb is a state, with accomplishment or achievement forms and causative forms derived morphologically from the state; similarly, (7b) illustrates the symmetrical case where the base form is a causative accomplishment, with accomplishment and state forms derived by application of morphological means.[1]

The lexical operation accounting for the causative/unaccusative alternation is formalized in Reinhart 2002 as in (8), involving suppression of one of the theta-roles of the original verbal grid, under interesting universal constraints formulated in terms of a thematic feature-system (cf. the discussion of (2) above):

(8) *Reduction*: V (θ_1, θ_2) $\rightarrow$ V (θ_2)

'Reduction' should be carefully kept apart from the lexical operation of 'passivization' underlying 'voice' alternations cross-linguistically, and essentially involving the existential interpretation of a lexically discharged theta-role, as shown in (9) below:

(9) *Saturation*: $\lambda x \lambda y$ R(x,y) $\rightarrow$ $\lambda y \exists x$ R(x,y)

On a similar vein, it has been argued that middle formation across languages involves application of a common lexical operation and a shared system of lexical representations based on thematic-feature decomposition (Marelj 2004). Significantly, it has also been proposed that the range of variation exhibited by middles cross-linguistically depends on the possibility of attaining the very same interpretive effects by application of syntactic means (Marelj 2004: Chapter 5). This way of accounting for linguistic variation within a common semantic core is highly reminiscent of the kind of syntactic pre-encoding of interpretive effects to be discussed in Section 4 below.

Let us now briefly consider the status of substantive universals in lexical semantics. In their most obvious form, these universals concern the presence of general constraints governing the building mechanism for lexical entries. One way of executing this research program consists in showing that the same function-argument structure applies indifferently across different semantic fields, ensuring a common cognitive base for superficially different lexical forms (Gruber 1965, Langacker 1987, Wierzbicka 1992). More precisely, the program can be executed by showing that (verb) lexical items are derived from a finite (and possibly very limited) array of words, hopefully constituting the exhaustive list of the primitive building blocks for a language lexicon (Wierzbicka's *Natural Semantics Metalanguage*). Alternatively, the program may be more indirectly implemented by means of generalizations across distinct semantic fields, by showing for instance that grammatical patterns used to describe physical objects in space are also used to describe non-spatial relations (Jackendoff, Langacker). A canonical example (adapted

[1] In fact, as observed by an anonymous reviewer, languages are also attested where the base form is an accomplishment/achievement. Clearly, this does not affect the logic of the argument made here, aimed to stress the pervasiness of lexical operations relating members of the *Aktionsart* classes under discussion.

from Jackendoff 2002) of a common abstract organization of superficially different semantic fields is shown in (10) below:

(10) *Spatial location and motion*

a. The messenger is in Istanbul	[location]
b. The messenger went from Paris to Istanbul	[change of location]
c. The gang kept the messenger in Istanbul	[caused stasis]

Possession

a. The money is Fred's	[possession]
b. The inheritance finally went to Fred	[change of possession]
c. Fred kept the money	[caused stasis]

Ascription of properties

a. The light is red	[simple property]
b. The light went from green to red	[change of property]
c. The cop kept the light red	[caused stasis]

An independent source of lexical universals concerns the way in which the internal structure of lexical items can be made more complex in order to account for the observation that certain classes of objects tend to belong, to a significant cross-linguistic extent, to more than one taxonomy, as in the canonical instance of co-predication involving 'book' (conceived both as an abstract information structure and as a physical object) in (11) below (adapted from Chomsky 2000):

(11) The book I have in mind, if I ever write it, will weigh more than 5 pounds

The relevant generalizations may concern both fine-grained descriptions of the lexical meaning of individual lexical items (Pustejovsky's *dot-objects*, exemplified with 'book' in (11); cf. Pustejovsky 1995) and fine-grained mechanisms of semantic composition, involving cases where it is necessary to access the internal semantic structure of words, as in the instance of verb-object composition in (12), whereby the activity selected by the verb (*reading* in the case of 'book' and *drinking* in the case of 'beer') corresponds to the use or function to which the two objects ('book' and 'beer') are canonically subjected (Pustejovsky's *'Direct Telic' quale*):

(12) a. I began/enjoyed the book (= I began/enjoyed reading the book)
b. I began/enjoyed the beer (= I began/enjoyed drinking the beer)

It goes without saying that these dynamic modalities of meaning description and meaning composition (Pustejovsky's *qualia structure*, Lieber's *body/skeleton* opposition; cf. also Melloni 2007) raise many intriguing questions concerning the conditions under which the encyclopedic information encapsulated in lexical items can be made accessible to the core of linguistic computations constituting the language faculty, both pre-syntactically (word formation and interpretation) and post-syntactically (phrasal interpretation). Clearly, understanding the universal principles and constraints governing the relation between conceptual structure and the feature-system underlying semantic computations is of tremendous importance for any theory of meaning aspiring to explanatory adequacy.

4 Universals and Phrasal Semantics

The strongest universalist position that one may take in phrasal semantics is expressed by the mapping principle in (13):

(13) **Universal**: The mapping between syntactic categories and semantic categories is invariant across languages

The mapping condition in (13) entails, for instance, that noun phrases (NPs) uniformly project into properties (type <e,t>), whereas determiner phrases (DPs) uniformly project into objects (type e) or generalized quantifiers (type <<e,t>,t>). The heuristic value of (13) clearly emerges from the observation that (13) is superficially disconfirmed by plenty of languages where determinerless nouns (NPs) are object-referring. This observation has actually led some researchers to adopt the position that the mapping of syntactic categories into notional categories should undergo parameterization, with NPs mapping into objects (and not into properties, as required by (13)) in languages featuring determinerless arguments, such as Chinese (Chierchia 1998). The only way to retain the universal validity of (13) consists then in arguing that superficially determinerless nouns, when they count as arguments, are necessarily introduced by a phonologically empty determiner. This heuristic strategy reveals itself to be extremely useful, since it allows one to show substantial empirical evidence in favor of the view that even bare nouns – and even in languages like Chinese – correspond in fact to DPs whenever used as arguments (see Longobardi 1994, 2006 for kind-referring bare nouns in Germanic; Cheng and Sybesma 1999, 2005 for bare nouns in Chinese; Engelhardt and Trugman 1998, Babyonyshev 1998 for bare nouns in Russian).

The result that argumenthood (and in particular object-reference) is closely connected to determiner phrases raises in turn the question concerning the conceptual link that should exist between determiner positions and reference to objects. A related observation is that only arguments have to be inflected for person, as shown by facts such as (14) below, where a (pro)nominal predicate expressing a property can only be resumed by a (default) third-person clitic (Longobardi 2006:18):

(14)	Se	Gianni	fosse	te	o	se	Maria	lo	fosse	/ *ti	fosse
	if	Gianni	were	you	or	if	Maria	CL3SgM	were	/ *CL2Sg	were

Since individuals are arguably denoted in the determiner position, and there is convincing evidence that person features are a prerogative of arguments, it is tempting to conclude that object-reference is made possible by associating the semantic content of Person to the determiner position. Pursuing this intriguing line of argumentation, Longobardi 2006 comes to the formulation of a universal semantic constraint linking object-reference to the semantic content of Person, along the lines of (15) below:

(15) Object reference is universally made possible by the Person feature

In the prototypical case of object-reference represented by proper names, association with person is realized by the expression of Person in the determiner

position and via the morphosyntactic forms of association of proper names with the determiner position studied in Longobardi 1994 (see also Delfitto 2002). For languages (such as Japanese) with no obvious manifestation of Person within morphosyntax, (15) predicts that object-reference be necessarily realized through free association of the semantic content of names with the semantic concept of Person as universally present within the conceptual system. Since the association of nouns with Person is not of a morphosyntactic nature in Japanese (or better, since there is no morphosyntactic pre-encoding of the semantic association of nouns with Person in Japanese), the syntax of object-reference will be radically simplified with respect to, say, Indo-European. More particularly: (i) the difference between common nouns and proper names is blurred (since there is no syntactic form of association of names with the determiner position, as in Longobardi's cases of (overt) N-raising); (ii) the difference between pronouns and names is blurred (since personal pronouns universally express a canonical form of realization of the person feature in the determiner position, but this strategy is entirely precluded to Japanese). If this analysis is on the right track, it may provide substantial conceptual and empirical corroboration for an important refinement of the universal in (13): the uniformity of mapping principles across languages is not limited to syntactic categories but can be tentatively extended to the set of formal features that we have consistently assumed, along this contribution, to represent the basic input of the syntactic computation. More precisely, (13) can be rephrased in the terms of (16):

(16) **Universal** (revised): The mapping between the grammatical features feeding linguistic computations and the notional categories universally realized within the conceptual system is invariant across languages

Given this framework of reference, the primary source of linguistic diversity is identified with the different degree of morphosyntactic pre-encoding attested in different languages (see especially Svenonius 2006; for a closely related view in the domain of temporal anaphora, see Giorgi and Pianesi 1997). It should be noted, however, that the associations realized in morphosyntax can be entirely realized, in principle, within the conceptual component. This fits thus quite well with the more general conclusion formulated above, according to which the basic ingredient of an adequate theory of meaning for natural language is a correct understanding of which parts of the Conceptual-Intentional systems are 'linguistically' activated, that is, activated for the purposes of the syntactic or semantic computations proper to the language faculty.

The universal in (16) is prima facie falsified by the observation that the ambiguity detected in (17) for English (and, more generally, Indo-European) is not found in other languages, where the reading corresponding in (18b) is expressed by means of 'shifted' second-person pronouns (i.e. 2nd person pronouns used anaphorically) (Schlenker 2003, Delfitto 2007):

(17) Bill told Hillary that she was too arrogant
(18) a. Bill told Hillary: "Hillary is too arrogant"
b. Bill told Hillary: "You are too arrogant"

This observation might lead one to propose that the 3rd person features realized on the pronoun 'she' in (17) are not assigned a universal interpretation, yielding a violation of (16). However, a possibility worth exploring is that the reading corresponding to (18b) arises as an effect of the non-interpretability of 3rd person features. Non-interpretable person features (technically, features that are added in the course of the syntactic computation as a result of a morphosyntactic operation of copying targeting an independent 3rd person antecedent) would then be exploited for the sole purpose of encoding a formal dependency between the pronoun 'she' and the antecedent 'Hillary' in (17). In this sense, 'she' would qualify as a 'minimal pronoun' (in the sense of Kratzer 2006). Suppose further that the formal dependency encoded by the non-interpretable 3rd person features realized on 'she' is interpreted by identifying the referent of the pronoun with the referent of the antecedent (co-valuation), crucially including the properties assigned to the antecedent in virtue of the theta-role that it fulfills (essentially, its being the 'addressee' within the speech act performed by Bill). Under this view (according to which pronominal anaphora involves sharing of thematically-related features), the direct speech reading corresponding to (18b), according to which Bill acknowledges that the person endowed with the property 'of being too arrogant' is the addressee of the speech act that he performs, would be expressed by the indirect speech construal in (18c), intuitively the correct empirical result:

(18) c. Bill told Hillary that Hillary, as the addressee of that very same speech act, was too arrogant

Now notice that if the reading in (18b) (corresponding to the use of 'shifted' 1st and 2nd person pronouns in non-Indo-European languages) can be discharged on the presence of non-interpretable 3rd person features on the pronoun, the possibility arises of preserving a universal interpretation of 3rd person features, *whenever they can be interpreted*. In particular, we may suggest that both in Indo-European and non-Indo-European languages, 3rd person is interpreted as expressing the presupposition that the individual referred to by the pronoun is distinct from the speaker and the hearer. In a nutshell, this means that the pronoun in (17) is simply translated into a variable ranging over semantic values distinct from speaker and hearer, along the lines of (19):

(19) Bill told Hillary that x was too arrogant

A natural possibility consists in interpreting x as co-valued with Hillary, giving rise to the reading corresponding to (18a) (accidental coreference, *de re* interpretation). The important fact to be noticed is that this reading arises universally, and corresponds to the unique interpretation that is available for *interpreted* 3rd person features. We conclude that the universal in (16) may actually be valid and constitutes in fact an important heuristic tool for the investigation of some intriguing interpretive properties of pronominal anaphora, intuitively related to *de re* / *de se* ambiguities.

5 Generalized Quantifiers and Set-Theoretic Universals

Starting from the seminal study in Barwise and Cooper 1981, it has become increasingly evident that a set-theoretic approach to the semantics of determiners (Generalized Quantifier Theory) may give rise to an interesting class of semantic universals. Given obvious space limitations, it is not possible to review here all the properties and constraints that have been proposed within this very productive stream of semantic research (for a first orientation, see Delfitto 1986 and the references cited therein). I will thus limit myself to discuss conservativity and logicality.

Consider first the universal in (20):

(20) **Universal**: All natural language determiners are conservative

Conservativity is formally defined in (21a) and gives rise to the kind of logical equivalences exemplified in (21b–c), for the determiners 'some' and 'most':

(21) a. *Conservativity*: D(A) is B $\leftrightarrow$ D(A) is A$\cap$B
b. Some student walks $\leftrightarrow$ Some student is a student who walks
c. Most students walk $\leftrightarrow$ Most students are students who walk

One of the interesting remarks to be made about conservativity is that it makes it possible to compute the different relations between two sets A and B expressed by different determiners without taking into considerations the elements of B that are not in A: in Barwise and Cooper's terminology, we say that a determiner D 'lives on' the set A (it goes without saying that this has non-trivial consequences for semantic parsing). Moreover, the striking fact about conservativity is that there are elementary set-theoretic relations that are not conservative (like those in (22) below) and that, significantly, cannot be expressed by a determiner in any known natural language. This may be taken to entail that conservativity is the core semantic constraint in the mapping between determiners and the notional categories associated to determiners.

(22) a. A = B is true iff A = A$\cap$B (identity)
b. <X,Y>: X>Y (Rescher's quantifier)

This conclusion is corroborated by the observation that potential counterexamples to conservativity concern elements that, under a closer scrutiny, are better assigned a categorial status different from determiner. This is the case for 'only-phrases', which do not license the sort of logical equivalences tied to (21a), as shown by (23) below:

(23) *Only Americans fight $\leftrightarrow$ Only Americans are Americans who fight

The fact that the purported equivalence is not valid is made clear by the observation that the right member of the bi-conditional in (23) qualifies as a tautology, since the truth-conditions for 'Only A is B' reduce to the requirement that B (the intersection between Americans and fighters) be included in A (the set of Americans). However, the point is that 'only', on distributional grounds, is not a determiner, since it can be

realized outside the nominal domain as a VP-modifier, as shown in (24) (it qualifies thus as an adverbial element; see Chierchia 1997):

(24) Americans only fight wars, they don't pursue peace ideals

This is not to say, of course, that the universal in (20) is not in need of some further qualifications (see for instance Westerståhl 1985). There is no doubt, however, that conservativity is a core property of the determiner system.

Let me now briefly consider logicality. The relevant question here is whether there exist syntactic categories in natural language whose notional counterpart can be characterized as a logical term. A logical term is a term whose interpretation is invariant under isomorphic structures defined on a given domain of interpretation. More particularly, a determiner is logic if it can be defined as a function from ordered pairs of natural numbers to truth-values. For instance, some of the most common English determiners can be defined as in (25) below, and qualify thus as logical ('a' corresponds to the difference between set A and set B, 'b' corresponds to the intersection between set A and set B):

(25) Every(<a,b>) = t if and only if a = 0 e b = n ($0 \leq n$)
No (<a,b>) = t if and only if a = n e b = 0
Most (<a,b>) = t if and only if b > a

It is thus tempting to think that DPs headed by a determiner qualify as logical categories:

(26) **Universal** (Logicality Thesis): Generalized quantifiers are universally the result of the combination of a logical determiner with a NP

Despite its intuitive appeal, we know by now that the alleged universal in (26) is false: there are DPs whose behavior is indistinguishable from the behavior of logical generalized quantifiers (in that for instance they pattern with logical DPs and not with names in licensing anaphoric relations) but that do not qualify as logical DPs on semantic grounds. A case in point is that of 'exception-phrases' like 'no students except five students in chemistry' (see Moltmann 1995 for a full discussion). Even negative conclusions such as the dismissal, on uncontroversial empirical grounds, of the potential universal in (26), can be of remarkable importance for a proper characterization of the mapping between syntactic and semantic categories. In particular, since generalized quantifiers are the most likely candidate for a logical status, the dismissal of (26) entails that none of the semantic categories that are associated to standard syntactic categories can be assigned a logical status. This shows in turn that the semantic computations that are part of the knowledge of natural language are not built up as a logic, that is, as a system intended to single out the set of valid sentences or the set of valid inferences.

6 Ontology and Grammar

Reference to events is a relatively well-established fact in natural language, since the seminal proposals in Davidson 1980 (see also Higginbotham 1985, Larson and Segal 1995).

The relation between semantic variability and ontological presuppositions gives rise to other intriguing research questions. Consider for instance the fact that determinerless nouns arguably denote kinds in a language such as English, whereas they simply introduce variables in Romance-like languages (see Longobardi 2001 for an updated discussion of this long-debated issue). One of the obvious questions that arise concerns the possible interplay between the ontological presuppositions carried by the semantics of natural language and the feature systems underlying the syntactic and semantic computations. It is often syntactic data that provide substantial evidence in favor of relatively subtle ontological distinctions that have been originally neglected. Here is a telling example. The sentences in (27) might be taken to suggest that both the *that*-clause in (27a) and the derived nominal in (27b) denote a proposition, given their perfect equivalence on interpretive grounds:

(27) a. That Bill arrived very late surprised Hillary
b. Bill's very late arrival surprised Hillary

However, notice that a *that*-clause is no longer freely interchangeable with a derived nominal when it counts as the selected complement of a predicate of propositional attitude. This is shown in (28):

(28) a. Hillary believes that Bill arrived very late
b. *Hillary believes Bill's very late arrival

The grammaticality contrast between (28a) and (28b) cannot be discharged on categorial selection properties of the predicate, since 'believe' is perfectly fine with other kinds of DP-complements, as shown in (28c):

(28) c. Hillary believes this

These observations seem to suggest that the explanation for the degraded status of (28b) resides in the fact that the objects referred to by *that*-clauses (propositions) are not the same as the objects referred to by derived nominals such as 'arrival'. Since *that*-clauses can also be used as arguments of psych-verbs, as in (27a), on a par with derived nominals, the conclusion might be that derived nominals are necessarily non-propositional, whereas *that*-clauses are ambiguous between a propositional and a non-propositional status (for a full discussion on this and other strictly related issues, see Zucchi 1993). Since there is evidence to the effect that complements of psych-verbs are non-eventive, we need to enrich the natural language ontology with entities endowed with an intermediate status between events and propositions (let us call them 'facts'; see Asher 1993 and Neale 2001).

Concerning reference to facts, a stimulating hypothesis is that the ontological commitment to facts in a language L may be subjected to grammatical encoding. This leads to unexpected relations between ontology and syntax. It can be shown

indeed that reference to entities that are non-eventive and non-propositional (arguably qualifying as 'facts') crucially depends on the presence of specific formal features that are interpreted in dedicated positions, and on the syntactic computations in which these features are involved. In Italian, we find for instance a singular contrast – in the possibility of anaphorically resuming a fact reported in a preceding utterance – between active and passive sentences, as can be seen in (29) and (30). A null pronoun can resume the fact expressed by the first sentence in the passive sentence in (30), whilst this is impossible in (29) (see Delfitto 2005 for a full discussion):

(29) L' uomo è autodistruttivo. *(pro) Dimostra che la natura umana ha aspetti negativi
the man is self-destructive it shows that the nature human has aspects negative
'Man is self-destructive. This shows that nature has a negative side'

(30) L' uomo è autodistruttivo. *(pro) È stato dimostrato dalla storia umana
the man is self-destructive it is been shown by the history human
'Man is self-destructive. This is shown by human history'

A natural hypothesis is that there are structural environments in which null pronominals get enriched with the interpretable feature required for fact-reference in the course of the syntactic derivation. In these contexts, null pronominals should be able to resume facts that have previously been introduced into the discourse domain. In this respect, consider the contrast below between (31) and (32):

(31) a. È scoppiata una bomba. *(pro) Mi ha riempito di sgomento
is exploded a bomb it CL1Sg has filled of dismay
'A bomb exploded. This fact frightened me'

b. È scoppiata una bomba. *(pro) Ha provocato enorme emozione
is exploded a bomb it has caused huge emotion
'A bomb exploded. This fact is caused great emotion'

(32) a. È scoppiata una bomba. (pro) È ormai noto a tutti
is exploded a bomb it is already known to all
'A bomb exploded. This fact is already known to everybody'

b. È scoppiata una bomba. (pro) È doloroso ma vero
is exploded a bomb it is painful but true
'A bomb exploded. This fact is painful but true'

c. È scoppiata una bomba. (pro) È un fatto
is exploded a bomb it is a fact
'A bomb exploded. This is a fact'

Syntactically, there is a clear difference between pronominal resumption in (31) and pronominal resumption in (32). In (31), the null pronominal is realized in the canonical spec-of-VP position proper to external arguments: in the course of the derivation, it never finds itself in a sisterhood relation with the selecting predicate. In (32), the null pronominal represents the subject of the small clause including an adjectival (32a–b) or a nominal (32c) predicate: in the course of the derivation, it finds

thus itself in a sisterhood relation with a fact-selecting predicate (*to be well-known*, *to be painful/true*, *to be a fact*). Arguably, it is this sisterhood relation between the fact-referring pronominal and the fact-selecting predicate that is responsible for the licensing of fact-reference in passive structures such as (30). Analogously, null pronominal subjects get computationally endowed with the interpretable feature required for fact-reference in the small-clause configuration proper to (32), before moving higher up for syntactic reasons.

If these considerations can be generalized across languages, ontological commitments in natural language are not only a question of arbitrary metaphysical choices or of logical regimentation. Rather, they universally manifest a systematic interplay with the feature-systems feeding the syntactic and semantic computations.

References

Asher, Nicholas. 1993. *Reference to abstract objects in discourse*. Dordrecht: Kluwer.

Babyonyshev, Maria. 1998. The possessive construction in Russian: A crosslinguistic perspective. *Journal of Slavic Linguistics* 5: 193–234.

Barwise, John, and Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4: 159–219.

Cheng, Lisa Lai-Shen, and Rint P.E. Sybesma. 1999. Bare and not-so-bare nouns and the Structure of NP. *Linguistic Inquiry* 30: 509–542.

Cheng, Lisa Lai-Shen, and Rint P.E. Sybesma. 2005. Classifiers in Four Varieties of Chinese. In *The Oxford handbook of comparative syntax*, eds. Guglielmo Cinque and Richard Kayne, 259–292. New York: Oxford University Press.

Chierchia, Gennaro. 1997. *Semantica*. Bologna: Il Mulino.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6: 339–405.

Chomsky, Noam. 1995. *The minimalist program*. Cambridge MA: MIT Press.

Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.

Davidson, Donald. 1980. *Essays on actions and events*. New York, Oxford: Oxford University Press.

Delfitto, Denis. 1986. Quantificatori generalizzati e universali semantici. *Rivista di grammatica generativa* 11: 65–98.

Delfitto, Denis. 2002. *Genericity in language*. Alessandria: Edizioni dell'Orso.

Delfitto, Denis. 2003. Results in semantics. *Lingue e Linguaggio* 2: 261–275.

Delfitto, Denis. 2005. On facts in the syntax and semantics of Italian. In *Romance languages and linguistic theory*, eds. Twan Geerts, Ivo van Ginneken, and Haike Jacobs, 15–35. Amsterdam: Benjamins.

Delfitto, Denis. 2007. A case-study on grammar, meaning and linguistic diversity: the nature of person features. In *Proceedings of the annual meeting of the Società Italiana di Glottologia – 2006*. Pisa: Scuola Normale Superiore.

Engelhardt, Miriam, and Helen Trugman. 1998. D as a source of adnominal genitive in Russian. In *Proceedings of the Sixth Annual Workshop on Formal Approaches to Slavic Linguistics*, eds. Željko Bošković, Steven Franks, and William Snyder, 114–133. Ann Arbor MI: Michigan Slavic Publications.

Fillmore, Charles J. 1982. Towards a descriptive framework for deixis. In *Speech, place and action*, eds. Robert J. Jarvella and Wolfgang Klein, 31–52. New York: Wiley.

Giorgi, Alessandra, and Fabio Pianesi. 1997. *Tense and aspect: From semantics to morphosyntax*. Oxford: Oxford University Press.

Gruber, Jeffrey Steven. 1965. *Studies in lexical relations*. PhD dissertation, MIT. Repr: Gruber, Jeffrey Steven. 1976. *Lexical structures in syntax and semantics*. Amsterdam: North-Holland.
Higginbotham, James. 1988. Is semantics necessary? *Proceedings of the Aristotelian Society* 88: 219–241.
Higginbotham, James. 1985. On semantics. *Linguistic inquiry* 16: 547–593.
Jackendoff, Ray. 2002. *Foundations of language*. Oxford: Oxford University Press.
Kratzer, Angelika. 2006. Minimal pronouns. Fake indexicals as windows into the properties of bound variable pronouns. Manuscript, University of Massachusetts, Amherst.
Ladusaw, William A. 1979. *Polarity sensitivity as inherent scope relations*. PhD dissertation, University of Texas at Austin.
Langacker, Ronald. 1987. *Foundations of cognitive grammar*. Stanford: Stanford University Press.
Larson, Richard, and Gabriel Segal. 1995. *Knowledge of meaning. An introduction to semantic theory*. Cambridge MA: MIT Press.
Levin, Beth. 1985. *Lexical semantics in review: An introduction*. Cambridge MA: MIT Center for Cognitive Science.
Levin, Beth, and Malka Rappaport Hovav. 1992. The lexical semantics of verbs of motion: The perspective from unaccusativity. In *Thematic structure: Its role in grammar*, ed. Iggy M. Roca, 247–269. Berlin: Foris.
Lieber, Rochelle. 2004. *Morphology and lexical semantics*. Chicago: The University of Chicago Press.
Longobardi, Giuseppe. 1994. Reference and proper names. *Linguistic Inquiry* 25: 609–667.
Longobardi, Giuseppe. 2001. How comparative is semantics? A unified parametric theory of bare nouns and proper names. *Natural Language Semantics* 9: 335–369.
Longobardi, Giuseppe. 2006. Reference to individuals, person, and the variety of mapping parameters. To appear in *Proceedings of Copenhagen Determiner Colloquium*, eds. Alex Klinge and Henrik H. Müller. Philadelphia, Amsterdam: John Benjamins.
Mairal, Ricardo, and Juana Gil. 2006. *Linguistic universals*. Cambridge: Cambridge University Press.
Marelj, Mariana. 2004. *Middles and argument structure across languages*. PhD dissertation, Uil-OTS, Utrecht university.
Melloni, Chiara. 2007. *Polysemy in word formation: The case of deverbal nominals*. PhD dissertation, University of Verona.
Moltmann, Friederike. 1995. Exception sentences and polyadic quantification. *Linguistics and Philosophy* 18: 223–280.
Neale, Stephen. 2001. *Facing facts*. Oxford: Oxford University Press.
Newmeyer, Frederick J. 2005. *Possible and probable languages. A generative perspective on linguistic typology*. Oxford: Oxford University Press.
Pinker, Steven. 2007. *The stuff of thought*. London: Allen Lane.
Pustejovsky, James. 1995. *The generative lexicon*. Cambridge MA: MIT Press.
Reinhart, Tanya. 2002. The Theta-system – An overview. In *Theoretical Linguistics* 28, ed. Wolfgang Sternefeld, 229–290. Berlin: Mouton.
Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26: 29–120.
Svenonius, Peter. 2006. Interpreting uninterpretable features. Manuscript, University of Tromsø.
Wierzbicka, Anna. 1992. *Semantics, culture and cognition: Universal human concepts in culture-specific configurations*. Oxford: Oxford University Press.
Wolff, Phillip. 2003. Direct Causation in the linguistic coding and individuation of causal events. *Cognition* 88: 1–48.
Van Valin, Robert D. 2006. Some universals of verb semantics. In *Linguistic universals*, eds. Ricardo Mairal and Juana Gil. 155–178. Cambridge: Cambridge University Press.
Vendler, Zeno. 1967. *Linguistics in philosophy*. Ithaca: Cornell University Press.
Westerståhl, Dag. 1985. Logical constants in quantifier languages. *Linguistics and Philosophy* 8: 387–413.
Zucchi, Sandro. 1993. *The language of propositions and events*. Studies in Linguistics and Philosophy, Kluwer Academic Publishers, Dordrecht.

The Evolution of Latin Word (Dis)order

Elisabetta Magni

Abstract The evolution from Latin to Romance languages involves a typological shift from SOV to SVO order where the change in Object position seems to be anticipated by a gradual construction-by-construction reorientation of constituent order. As a matter of fact, since Early Latin coexisting patterns determine a kaleidoscopic surface complexity that makes it difficult to reduce this language to a coherent system. On the assumption that the typology of word order flexibility is closely intertwined with the diachrony of word order change, this paper investigates the factors influencing grammatical variation and change by discussing both the sources that give rise to the various constructions and the mechanisms governing the choice between alternative orders at different stages. The results will show that, in some cases, grammatical variation depends on processes that are partially independent of the OV/VO dichotomy, and that typological regularities and irregularities in word order typology can be diachronically motivated.

Keywords Typology · diachrony · language change · word order · Latin · adpositions · adjective · genitive construction · comparative

1 Introduction[1]

Discussing Greenbergian proposals for the so-called dynamic comparison or diachronic typology,[2] Bybee affirms that: "the true universals of language are the mechanisms of change that propel the constant creation and re-creation of grammar" (Bybee 2006: 179–180). As far as Latin is concerned, re-creation of grammar features a huge structural reorganization, where the typological shift from SOV to SVO order still presents puzzling issues and not fully understood phenomena.

E. Magni
University of Bologna

[1] I would like to thank Alberto Nocentini, Sergio Scalise, Carlotta Viti and, above all, Edoardo Vineis for their valuable comments on earlier drafts of this paper, with the customary exculpations.

[2] Developed in Greenberg (1969).

S. Scalise et al. (eds.), *Universals of Language Today*, 225–251,
DOI: 10.1007/978-1-4020-8825-4_12,

The evolution from Latin to Romance languages involves three major facts: the emergence of a fixed, more rigid word order, the change from synthetic to analytical structures, and the erosion of nominal morphology. Many attempts have already been made to account for the complex interplay of these phenomena and the overall change (Vennemann 1974, Sasse 1977), and countless studies deeply investigated the ordering of specific structures (Marouzeau 1922 and 1949, Linde 1923), more recently focusing on pragmatic (Panhuis 1984, Del Vecchio 1989, Devine and Stephens 2006) and typological approaches (Adams 1976, Baldi 1979, Lehmann 1979 and 1991, Elerick 1991, Bauer 1995, Magni 2000, Nocentini 2005).

In this paper I will endeavor to explore this broad issue by combining the rich data collection of traditional philological approaches with the more explanatory insights of modern theoretical linguistics, and by interfacing syntactic analysis with semantic and pragmatic interpretations, in accordance with a functionalist approach. More specifically, the investigation will concentrate on early cracks in the SOV system,[3] and on the gradual construction-by-construction reorientation of constituent order as reflected in coexisting patterns from Early Latin onward. For present purposes, I shall limit myself to a closer examination of adpositions, coordinative, adjectival, genitival and comparative constructions. Limitations of space prevent me from discussing here the relative clause, but some considerations in this regard will be found in the conclusions.

2 Consistency and Change

2.1 Traditional grammars of Latin inform that, in spite of the variety allowed by a full-fledged inflectional system, word order was not indiscriminately 'free', and the distinction between '*traditionelle*' and '*okkasionelle Wortstellung*'[4] is repeatedly mentioned (Kühner and Stegmann 1976 vol. 2: 589–590, Hofmann and Szantyr 1965: 397–398). In fact, the rigorous analysis developed in Marouzeau's monumental work offers much valuable illustration to the introductory claim that Latin word order is free but not 'indifferent', in the sense that, generally, although every term in the sentence lacks a fixed position, two different word orders are not synonymous.

> l'ordre des mots en latin est libre, il n'est pas indifférent. Libre en ce sens que, sauf exception, il n'y a pas pour chaque terme de la phrase une place attitrée, obligatoire. Mais non pas indifférent, parce qu'en general deux ordres possibles ne sont pas synonymes.
>
> (Marouzeau 1922: 1)

More recently, after a broad investigation of the so-called 'communicative dynamism' governing constituent order, Panhuis affirms that "[f]rom a syntactic point of view, word order in Latin is indeed almost free. But pragmatically, it is not free, but organized in a communicative perspective" (Panhuis 1984: 156).

---

[3] Which for some scholars are also symptomatic of the delay of the verbal phrase with respect to the noun phrase (Bauer 1995, Adams 1976).

[4] That is, between normal (or habitual) and occasional word order.

Since pragmatic (and other) factors are often assumed to influence the choice of constituent order, it becomes evident that the term 'free' is misleading,[5] and that Latin can be more appropriately ascribed to a subtype of flexible order languages: those in which one pattern is most frequent and can be assumed as the unmarked, neutral order (Dryer 2008a).

According to Quintilian there was, indeed, a preferred order, namely verb final:[6]

> *verbo sensum cludere multo, si compositio patiatur, optimo est: in verbis enim sermonis vis est*
> 'It is by far best, if composition allows it, to close the sense with the verb, for the force of language lies in verbs'
>
> (Quint. *Inst*. 9, 4, 26)

As a general rationale, it has been argued that languages with flexible order tend to be OV as basic type, and this is also consistent with the Universal 41,[7] because languages with a case system have much freer word order possibilities than languages that rely on order for marking grammatical functions. In connection with this, we may thus follow the widely accepted hypothesis that, starting from PIE as an OV language, Early Latin is likely to have been an OV language as well.

2.2 The typological consistency of PIE, however, has been the object of ongoing debate, and the same discussion also applied to Classical Latin,[8] where the attested patterns are quite intricate and, at first blush, inconsistent with any specific dominant order. By way of illustration, consider the following statement:

> By the time of Classical Greek and Latin, the OV syntactic pattern of PIE had been largely modified to a VO pattern. Yet the VO pattern was still inconsistent in the early classical period of both Greek and Latin, as is illustrated by such relic patterns as OV comparatives, and also in Classical Sanskrit. In spite of such inconsistencies, this stage of development is that 'reconstructed' by comparison and represented for the parent language in the standard handbooks, such as Brugmann's.
>
> (Lehmann 1974: 238)

Actually, pure types without disharmonic structures are rare, and many languages exhibit more than one order for at least some pairs of elements, or more flexible order for some elements, and less flexible for others.

---

[5] I will also avoid here both the label 'non-configurational' and the term 'scrambling', which is adopted, for instance, in Bolkenstein (2001) and in Devine and Stephens (2006).

[6] As can be seen from the data summarized in Hofmann and Szantyr, which repeat the percentages given in Linde (1923), the verb in clause-final position was particularly frequent in the sober and matter-of-fact language of professional authors and historians: "Die nüchtern-sachliche Sprache der Fachschriftsteller (Cato, Gaius) und der Historiker (Ceas., Sall., Liv., Tac., auch noch Iust. al) bevorzugt im Hauptsatz, mehr noch im Nebensatz die ESt [*sc*. Endstellung]", Hofmann and Szantyr (1965: 403).

[7] That says: "If in a language the verb follows both the nominal subject and nominal object as the dominant order, the language almost always has a case system".

[8] For a thorough discussion on PIE see Lehmann (1972), (1974) and Miller (1975), further observations and bibliography can also be found in Baldi (1979) and Nocentini (1992). With regard to Latin, see the analysis and the data in Adams (1976) and Panhuis (1984).

In a synchronic perspective, a closer examination of each alternating pattern may help on deciding how to classify a language according to the characteristic at issue. But when analyzing languages without living speech communities and whose documents are often stylized,[9] this task is further complicated, because in some contexts it is splitting hairs to attempt to distinguish whether an order is pragmatically neutral or has some added pragmatic effect,[10] and judgments about markedness tend often to reduce to simple frequency counts.

Under these circumstances, in my opinion, it would be preferable to abstain from speculating about an overall sketch based on 'marked' vs. 'unmarked' patterns,[11] and to concentrate the efforts upon the evolutionary paths that motivate coexisting constructions and grammatical variation.

The historical perspective adopted here agrees with Givón's assumption that both regularities and irregularities in word order can be diachronically motivated. On this interpretation, it is also evident that the typology of word order flexibility is closely intertwined with word order change, because "a period of word order flexibility is often part and parcel of the diachronic mechanisms via which one rigid word order changes into another" (Givón 2001: 271).

The kaleidoscopic surface complexity of Latin can thus be viewed as the consequence of the fact that multiple factors governed the choice of constituent order at different stages, and that the reordering took place at different periods in the various structures.

On the basis of these presuppositions, there are now two clearly defined tasks to be carried out. The first task involves researching the sources that originated alternating constructions. The second task entails a thorough investigation of the processing factors influencing grammatical variation and language change.

3 Adpositions

3.1 Adpositions are a thorny issue to start with, since Latin has both prepositions and postpositions. One of the basic tenets of typology is that the correlation between OV order and postpositions described in Universal 4 is a strong tendency[12] that can be stated by means of a bidirectional implicational universal: 'If a language is OV, then it is postpositional, and if it is postpositional, then it is OV' (Dryer 2007a: 89).

---

[9] The poetic, formalizing, or archaic style of many literary texts obviously mischaracterizes the actual daily use of the language, which might have exhibited a different or more regular syntax than written legacy indicates. In this respect, another limitation that has to be taken into account is that documents are "transmitted to us by copyists who can have an unhelpful tendency to 'normalize' word order" (Devine and Stephens 2006: 5).

[10] For a critical survey of the role of pragmatic factors, see Del Vecchio (1989), for a more recent application and analysis, see Devine and Stephens (2006).

[11] As in the formalist approach adopted in Giusti and Oniga (2006).

[12] Indeed so strong that Friedrich (1975: 39) writes: "the preponderance of prepositions in early Indo-European languages decreases the likelihood of a dominant SOV order".

In fact, according to the WALS data (Dryer 2008b), OV languages with prepositions, such as Kurdish, are the least frequent type, and Dryer also observes that if a language changes the order of one of these two features, it is almost always the order of object and verb that changes first. But it seems to go the other way round in Latin, where postpositions have become prepositions when the order was still predominantly OV.

Following a widely accepted reconstruction (Lehmann 1974, Cuzzolin, Putzu and Ramat 2006), most of the preverbs and adpositions in Latin can be taken back to PIE particles or adverbial elements involving local or temporal notions, and with rather free movement characteristics. In their unmarked preverbal position, these forms could either became fused with the following verb, thus originating bound preverbs, or attach to the preceding noun, thus forming postpositions,[13] which served first as a sort of adverbial accompaniment, independently of the cases, and then arrived to 'govern' the noun by specifying its case, and eventually replacing oblique cases.

3.2 The process of transition from postpositions to prepositions is already attested in the Sabellic[14] (or Osco-Umbrian) languages, which are somewhat archaic in general and may be expected to give useful information on the first stages of Latin, to which they are closely related. Umbrian, for instance, still has five genuine postpositions, namely *-ař, -en*, *co(m)*, *-per, -per(t)*:[15] e.g. *asamař* 'to the altar', *erucom* 'with him' (other examples in (7b) and (11a)). Similar forms and usages are occasionally found in Oscan as well: the phrase in (1)[16] shows an interesting occurrence of the ending *ín*, which is a fusion of loc. *-ei* with the postpositive *-en*, extended to the adjective.[17]

---

[13] As a matter of fact, in some of the oldest examples it is often difficult to decide whether the form belongs to the noun or to the verb (Baldi 1979: 54). The main support for a postpositional view of PIE comes from Hittite, which was exclusively postpositional, and this was also the general rule for Vedic, though preposing was frequently found as well. Other evidence can be found in relic forms of Greek and Latin, but for a thorough discussion on this controversial matter see Baldi (1979: 51–53).

[14] For the use of the label 'Sabellic', instead of 'Osco-Umbrian', or the partly ambiguous (at least in Italian usage and studies) 'Italic', and for a thumbnail sketch of these languages, see Rix (2003).

[15] The latter has restricted uses, mainly after numerals, to form adverbs like U. *trioper, triiuper* 'three times', O. *petiropert* 'four times', cf. Lat. *semper* 'always', *paulisper* 'for a while', *parumper* 'for a little while'. With the meanings 'beyond, past, after' the form is prepositional: *pert spinia* 'past the column'. Umbrian also has the suffix *-to* 'from', whose problematic status and uncertain origin are discussed in Nocentini (1992: 218).

[16] Interlinear glosses conform to the notational conventions of the Leipzig Glossing Rules (see: http://www.eva.mpg.de/lingua/files/morpheme.html).

[17] In Umbrian the combination of the locative ending *-ē* with the postposition, yielding a form in *-en* or *-em* was very frequent, and its extension to the adjective forms "was probably favored by the parallelism between Locatives with and without *m* and Accusatives with and without *m*, where the m of course appears in the adjective also" (Buck 1904: 114).

(1) *húrtín* *kerríín*
garden.in:LOC.SG of.Ceres.in:LOC.SG
'in the garden of Ceres'
(*Tabula Agnonensis*,[18] a 1)

3.3 The Latin cognates of these forms are normally preposed. On the whole, the place of adpositions was moderately free, and Marouzeau (1953: 67) suggests that discontinuous structures like the one in (2) still reflect speakers' hesitation between preposing and postposing:

(2) *arbusta* *per* *alta*
timber-tree:ACC.PL through tall:ACC.PL
'through tall timber-trees'
(Enn. *Ann*. 187)

Although we need to be cautious when considering the language of poetry, it is beyond doubt that prepositions were frequently separated from their case by an attributive adjective or its equivalent, or by other modifiers of the case. Moreover, inserted elements like enclitics were always possible in fixed expressions and, especially with *per*, in adjurations and exclamations. This can be seen from the examples in (3):

(3) a. *contra=que* *legem*
against=and law:ACC.SG
'and against the law'
(Cic. *Verr*. 3, 92)

b. *per* *ego* *te* *deos* *oro*
by I:NOM you.ACC god:ACC.PL pray:PRES.1SG
'I pray thee, by the gods'
(Ter. *And*. 834)

Instances of postposing in Latin are usually accounted for either as a stylistic strategy (anastrophe), or as the residue of the former organization occurring in specific contexts. In this respect, *cum* has been longest in use as postposition, always following personal pronouns in fixed expressions like *mecum*, *tecum* 'with me, you, etc.', and frequently following relative pronouns as well, as in *quocum, quibuscum* 'with whom' (cf. Germ. *womit*), from the Classical period on, however, these usages already coexist with the corresponding prepositional structures (Ernout and Thomas 1951: 101).

Other monosyllabic adpositions developed more quickly and their postposition is even more restricted. The form *de*, for instance, follows the relative pronoun only in juridical formulae such as *quo de agitur* 'the point in question', and in Early Latin position after the relative is found here and there also with

[18] The dedicatory tablet of Agnone, a small bronze tablet inscribed on both sides, was found near Isernia (Molise) in 1848 and dates to about 250 B.C.E.

ex 'from', *ad* 'to', *ab* 'from', *per* 'through', *pro* 'in front of': cf. *quoad* 'as long as'. Disyllabic adpositions, on the other hand, are easier postposed, especially after a relative or a demonstrative: *quocirca* 'wherefore', or *quapropter*, *eapropter* 'therefore', etc.

The frequent usage in fixed expressions, however, triggers erosion, as in the example (1), and semantic bleaching, which seems to explain the hypercharacterized constructions of the early inscriptions in (4) (Marouzeau 1949: 46):

(4) a. *cum quicu* 'with whom'
(CIL XI, 5779)

b. *cum qua com* 'with whom'
(CIL VI, 164, 14)

For Baldi, the coming into being of prepositions takes place via reanalysis of contexts like the one in (5):

(5) [. . .] *eas=que* *ad* *urbem* *adducere*
that:ACC.PL.F=and to city:ACC.SG lead:PRES.INF
'and to lead them [legions] to the city'
(Cic. *Fam*. 12, 23, 2)

Here, he says, "preverbs are copied from their preverbal position into prenominal position" (Baldi 1979: 57), supplementing nouns with the same addition of notions (local, temporal, etc.) by which they modify verbs, in order to convey those syntactic and semantic details that the general reduction of PIE cases was making difficult to express. Under this view, the rise of the prepositions "in fact has nothing whatever to do with word order or word order change" (Baldi 1979: 53).

Nonetheless, the usage illustrated in (5) is relatively recent and rare (Nocentini 1992: 228). Moreover, Baldi's explanation subordinates the emergence of prepositions to the specialization of the preverbs, but these phenomena, as we will see in a moment, are not necessarily interdependent.

3.4 Adpositions from grammaticalization processes involving head nouns in genitive constructions, on the other hand, are more consistent. Comparable forms already existed in Sabellic languages, where we find O. *amnud* and U. *paca* 'for the sake of': an example from Umbrian is given in (6):

(6) *ocrer* *pehaner* *paca* = Lat. *arcis piandae causa*
citadel:GEN.SG purify:GDV.SG for the sake of
'to purify the citadel'
(*Tabulae Iguvinae*,[19] VI a, 20)

[19] The seven bronze tables found at Gubbio (ancient Iguvium, 25 km NE of Perugia) in 1444 are mostly written in the Umbrian script (I-IV and Va-Vb7), but partly also in the Latin alphabet (Vb7-18 and VI-VII). The former date to the beginning of the third century B.C.E., the latter to the first half of the first century B.C.E.

Here the word *paca*[20] has the same functions of Latin forms like *causā, gratiā* 'for the sake of', etc., which in traditional grammars are also called 'improper prepositions' (Buck 1904: 211), or 'prépositions postposées' (Ernout and Thomas 1951: 99–101). These are generally relational nouns 'frozen' in the ablative form, and partially grammaticalized as postpositions.

Assuming that the prepositional phrase was definitely right branching since Early Latin, Bauer (1995: 135) explains their postposing as a mean to distinguish them from the regular nouns.[21] As we have seen, however, adpositions do not constitute a homogeneous category, neither in synchrony nor in diachrony.

Alternatively, we may thus hypothesize a fluid situation, where adpositions from different sources evolve at different rates and where each kind "grammaticalizes on its own, following its own diachronic trajectory that may or may not be determined by clausal (OV vs. VO) syntax" (Givón 2001: 245).

More precisely, the forms of nominal origin, which reflect a stage where the structure GN was unmarked (Nocentini 1992: 226), remain more or less rigidly postposed until the change to SVO is completed, whereas adpositions from particles and adverbial elements retain longer their free and variable status.

3.5 As we have seen in 2.1, the systematic relationship between preverbs and adpositions is traditionally accounted for in terms of two different reanalyses of the basic structure NP + X + V, where X represents the oldest category Adv/Adp/Prev (Cuzzolin, Putzu and Ramat 2006), with still undifferentiated functions. That is, the genesis of bound preverbs presupposes a reanalysis NP + [X + V], whereas a reanalysis [NP + X] + V explains the rise of postpositions.

As Nocentini (1992: 229) points out, however, the PIE stage theoretically admits six possible constituent orders, as follows:

1) NP + X + V	3) V + X + NP	5) *NP + V + X
2) X + V + NP	4) X + NP + V	6) *V + NP + X

Since asterisked sequences are rarely attested indeed, it becomes also evident that, as particles or adverbial elements specialize as preverbs or adpositions, two out of four remaining orders are suited to originate prepositions (namely 3 and 4); the second one can easily favor the emergence of preverbs, and only the first one is the possible source for both preverbs and postpositions.

On this interpretation, statistics are unfavorable to the latter outcome and, on the other hand, the development of prepositions can be considered as independent from word order change. In consonance with this, the role of adpositions with regard to

---

[20] This form is the ablative singular from **pākā-* '*pact*', cf. Lat. *pāctum*, *pacīscor, pāx*.

[21] "The ambivalence of the form, which was at the same time an adposition and a noun, might explain why this type of phrase did not change into an RB [*sc*. Right Branching] structure. Being postposed the adposition was thus distinct from the regular noun" (Bauer 1995: 135).

the typological description of the Indo-European languages appears to be strongly reduced (Nocentini 1992: 229).

4 Coordinative Constructions

4.1 The role of particles in coordinative constructions offers other useful insights about coexisting patterns. According to Stassen (2000), postposed copulative conjunctions of the type 'X Y *-and*' correlate with OV order, whereas VO order correlates with preposed coordinators of the type 'X *and* Y'. The coexistence of two constructions is a typological rarity of some IE languages that, beside inherited enclitic forms derived from PIE indefinite $*k^we$, display new preposed coordinators from locative and temporal particles, so that we have Vedic *ca* vs. *utá*, Ancient Greek τε vs. καί, and Latin *-que* vs. *et*. [22]

The specific functions of the first kind of coordinators have been described by various scholars:

> $*k^we$ seems to have been a means of indicating complementary unity, that is to say: it was a marker pointing to, or emphasizing, the fact that two (or more) words of the same category (substantives, adjectives etc.) were not only considered as belonging together, but constituted a complementary pair (or set). Hence the predilection for the double $*k^we$ and the frequent occurrence of 'opposites' connected by this particle. These 'opposites' were really complements.
>
> (Gonda 1954: 189)

> Il semble que le type le plus ancien soit représenté par τε . . . τε dans des expressions volontiers formulaires, qui associent des *objects*, ou des *êtres*, ou des *qualités* qui sont considérés comme naturellement couplés.
>
> (Humbert 1960: 435[23])

> [Latin *-que*] servant surtout à unir des mots formant couple: *domi bellique*, *senatus populusque Romanus*, - ou de sens voisin: *vis amicitiae concordiaeque*
>
> (Ernout and Thomas 1951: 372[24])

As Viti (2006) points out, the preceding statements suggest that in Vedic, Ancient Greek and Latin, the recessive pattern is related to natural coordination, whereas the new pattern is related to accidental coordination, a typological distinction recently investigated by Wälchli (2005).

What is relevant here is that the contrast between the two constructions with *-que* and *et* goes beyond syntax, and involves a functional competition based also on pragmatic and semantic factors. In fact, the use of forms etymologically denoting

---

[22] As it is well known, the Latin form *et* is etymologically correlated to Vedic *áti* 'beyond' and to Ancient Greek ἔτι 'further, moreover'.

[23] 'It seems that the oldest type is represented by τε . . . τε in expressions often formulaic, which associate objects or beings, or qualities that are regarded as coupled by nature'.

[24] '[Latin *-que*] mostly used to join words forming a couple: *domi bellique*, *senatus populusque Romanus*, - or with similar meaning: *vis amicitiae concordiaeque*'.

an indefinite union for natural coordination, and the use of locative and temporal particles for accidental coordination entail different methods of textual cohesion.

The natural coordination of noun phrases is a static kind of linkage, where the mention of one conjunct makes the other conjunct predictable, and where both conjuncts represent types rather than tokens of the referent, as can be seen from typical usages like *di atque homines* 'gods and men' (Pl. *Ps*. 381 and 600), *dulcia atque amara* 'the sweets and the bitters' (Pl. *Ps*. 694), etc. Moreover, that the conjuncts of a *k^we-coordination are typically neither modified nor specified is due to a generic presentation, which is compatible with the indefinite function of the pronominal stem *k^wi-/k^wo-.

On the contrary, *et* favors modification and specification, which anchor the referent in the current discourse and represent typical devices of conversational cohesion, as illustrated in *argentum* [...] *et obsignatum symbolum* 'the money ... and the token with the impression' (Pl. *Ps*. 1091–1092) or in *symbolum* [...] *et hoc argentum* 'the token ... and this money' (Pl. *Ps*. 598), where the conjuncts are also separated by other constituents. On the whole, accidental coordination is dynamic in nature, as its conjuncts are in principle unpredictable and, in the clause domain, convey foregrounded information (Viti 2006).

After a period of functional overlap, the difference in meaning between the two coordinators fades away, and the form devoted to accidental coordination generalizes at the expense of the form expressing natural coordination. All of this leads us to the conclusion that here, as in other cases, patterns from different sources coexist for a long time, because pragmatic, semantic and syntactic motivations make them suitable for different purposes and usages.

5 Adjectives

5.1 The ordering of modifying adjectives illustrates another case where grammatical variation depends on mechanisms that are partially independent of the OV/VO dichotomy.

As shown by Dryer (1988, 1992), the relation between adjective and noun is a noncorrelation pair, since the NA order is more common than AN order, both among OV and VO languages. In this regard, an additional problem is that in many languages of the world adjectives do not form a distinct word class. In the Indo-European domain the absence of a clear-cut distinction between nouns and adjectives gets reflected in the treatment of the parts of speech operated by the ancient grammarians, and the similarities between these lexical categories are manifested by the occurrence of several shared properties, such as: taking the same set of inflectional affixes for gender, number, and case; having the same set of derivational affixes; denoting a property when occurring in the adnominal position, and the possessor of this property when occurring in the head-noun position (Bhat 1994: 165).

Moreover, the inflectional marking of adjectives is correlated with a relative independence, rather than dependence in noun phrases, and so adjectives are allowed

to be shifted in different positions, and to form loose paratactic structures where items from the same category are juxtaposed and the modifier-modified distinction is left unspecified (Bhat 1994:170–171).

5.2 As far as Latin is concerned, similar observations apply to the status of the nominal group that, at a general level "is not integrated tightly enough to be called a phrase. There is neither tight syntactic cohesion in such syntagms nor a pronounced hierarchical inequality between the substantival head and the various kind of modifiers" (Lehmann 1991: 229). Consequently, the Latin adjective does not form a phrase with its head noun, but rather is coreferential with it.[25]

Given this peculiar situation, it is hardly surprising that the rules for the position of adjectives are so complicated, and also changeable through time. In his description of the Indo-European adjective Delbrück (1900: 94–100) puts forward a preliminary divide between simple/prenominal and derived/postnominal adjectives, then concluding that they normally precede the noun, but predicative use and emphasis determine a 'deviant order' in many languages.

On the contrary, Hofmann and Szantyr (1965: 406) simply state that "Im Idg. scheint das Adj. habituell nachgestellt worden zu sein",[26] and this behavior is actually attested in Sabellic languages, where "the adjective regularly follows its noun, but may precede it if emphatic" (Buck 1904: 224). Curiously enough, however, all the examples that he provides for the AN order are similar to the phrase in (7b):

(7) a. *ceus* *Bantins*
citizen:NOM.SG of.Bantia:NOM.SG
'citizen of Bantia'

(*Tabula Bantina*,[27] 4, 19)

b. *destruco* *persi*
right.at:ABL.SG foot:ABL.SG
'at the right foot'[28]

(*Tabulae Iguvinae*, VI b, 24 and 38)

For Adams these usages are in perfect keeping with those of Latin, where anteposition "was reserved for adjectives which were emphatic or which carried a strong emotional content" and where, as a general rule[29]:

---

[25] Not surprisingly, in Pāṇini's *Aṣṭādhyāyī* qualifier (*viśeṣaṇa*) and qualified (*viśeṣya*) are simply defined as *samānādhikaraṇa*, which basically means 'coreferential'; accordingly, their relationship is independent from constituent order and flexibly interpretable on the basis of contextual factors (Bhat 1994: 170–171).

[26] "In Indo-European the adjective seems to have been usually postposed"

[27] Discovered in 1793 in Apulia, this bronze tablet contains a series of municipal regulations for the town of Bantia (S. Maria di Banzi), and dates approximately to the last quarter of the second century B.C.E.

[28] In Umbrian the ablative with this postposition develops "a distinctly locative sense" (Buck 1904: 203).

[29] In Hofmann and Szantyr (1965: 406) the rule sounds as follows: "Als Regel darf gelten: objectiv bestimmende (intellektuell-sachliche) Adj. stehen habituell nach (*ius civile* wie *ius civium*, *populus*

> objective adjectives regularly follow the noun at all periods. The adjectives which *par excellence* belong to this category are those derived from proper names [...] and the possessives *meus, tuus, suus*, etc. On the other hand subjective adjectives - adjectives expressing a personal judgment, affective or emotive words (e.g. *bonus, malus, iustus, improbus*, etc. - usually precede the nouns.
>
> (Adams 1976: 88–89)

To be precise, this alleged distinction between objectivity and subjectivity entails a semantic and functional contrast that relates both to different (and variously labeled) adjective categories and to their usages in different contexts.

As in other languages like Spanish, French and Italian, Latin adjectives can either precede or follow the head-noun also coding the difference between 'literal' and 'figurative' sense. This can be seen from the couples in (8a–b)

(8) a. *dulces voluptates* 'sweet-flavored pleasures'
(Cic. *Fin*. 2, 30)

aqua dulcis 'fresh water'
(Cat. *Agr*. 106, 1)

b. *herba sanguinaria* 'herb that stanches blood'
(Col. *R.R.* 7, 5, 9)

sanguinaria iuventus 'blood-thirsty young men'
(Cic. *Att*. 2, 7, 3)

According to Bauer (1995: 73), these examples are consistent with the rules summarized in (9):

(9) [[Descriptive Adj.] Noun] in unmarked order;
NA in marked order

[Noun [Distinctive Adj.]] in unmarked order;
AN in marked order

She also claims that this situation reveals a diachronic change affecting first distinctive adjectives,[30] whereas the subcategory of descriptive adjectives remained left branched over time, but her explanation seems to entail a circular reasoning: "[t]he preposing of a subcategory of Latin adjectives can be considered an archaic feature and can be explained by its non-distinctive value" (Bauer 1995: 73).

5.3 In my opinion, a better explanation for this different behavior can be found in the very nature of the so-called distinctive adjectives, which are often derived from nouns (see Delbrück's divide mentioned in 5.2) and, in contrast with descriptive adjectives, do not display comparative and superlative forms because they do not

---

Romanus, *navis longa* als Schiffstyp usw.), qualifizierende (affektische) treten unter der Wirkung der Betonung voran (qualitätsadj. wie *bonus*, *pulcher*, Intensitätsbezeichnungen wie *magnus*, *summus*, *ingens* u.ä.). For more data see' also the detailed account in Marouzeau 1922.

[30] As it is well known, this distinction goes back to Marouzeau (1922), whose terms *adjectif distinctif* and *adjectif descriptif* have no equivalent in English grammatical terminology. Also accepted by Touratier (1991), the divide roughly corresponds to the distinction between determiners and qualifying adjectives.

focus on a single property but rather on the cluster of properties suggested by their general meaning 'related or pertaining to X', where X is the base. By contributing to participant-identification in their classifying function,[31] they are closer to nouns, more likely to recategorize as nouns[32] and, above all, are more independent than descriptive adjectives.

Following Lehmann, we may thus say that, at the general level of the attributive relationship, position relates to bondedness.[33] More specifically, prenominal position correlates with a tighter syntactic bond between attribute and head noun: in the case of adjectives this corresponds to the expression of properties that are inherent and essential, and to the functions of description and concept formation. Postnominal position, on the other hand, makes for a looser syntactic bond that involves the expression of accidental properties, more suited to discriminate an entity as against others, and to object identification (Lehmann 1991: 222–223).

From the above premise it can be easily explained why adjectives with prototypical qualifying function preferably take the prenominal position, which is incidentally more consistent with OV syntax, but has in fact little to do with typological correlations or emphasis, being rather used to mirror the higher level of dependency of the attribute. Within the Latin nominal group, however, the relative independence of the *nomen adjectivum* vis-à-vis its head noun, is normally in keeping with the postnominal position, which thus appears as the unmarked option for most adjectives. And this is particularly true for distinctive adjectives, which form structures that are more 'appositive' in nature, and also entail a sort of 'possessive' relationship, like the usually postposed possessives *meus, tuus,* etc.

This seems to explain why the NA order seems to emerge first in ancient religious formulae and expressions like those in (10):

(10) *campus Martius* 'field of Mars'
virgo Vestalis 'vergin Vestal, priestess of Vesta'
flamen Dialis 'flamen, high priest of Jupiter'
collis Quirinalis 'Quirinal hill, hill of Quirinus'
sacerdos Veneria 'priestess of Venus'

The phrases above also exhibit some repeatedly observed parallelisms with those genitive constructions where, according to Adams, "the names of gods were highlighted by postposition" (Adams 1976: 89).

---

[31] As de Carvalho (1991: 255) clearly states, "C'est donc *conjointement* mais non *solidairement* que, dans un groupe nominal ainsi formé, adjectif et substantif parteciperaient à la représentation du réel dont ils livrent, chacun de son côté, des aspects partiels, de statut inegal, mais complémentaires, et jugés également indispensables".

[32] See for instance the discussion on the formations in *-arius* in Magni, forth.

[33] According to Lehmann, this generalization, which is based on cross-linguistic studies of form and meaning correlations in complementation (e.g. the 'binding hierarchy' in Givón 1980), is "true for modifiers in general, not only for adjective attributes" (Lehmann 1991: 223); see also the observations on the noun phrase in Foley (1980).

6 Genitive Construction

6.1 The genitive construction, which in Latin marks the possessor, covers a wide range of meanings, such as kinship and part-whole relations, possession or ownership, and various abstract relations.

According to Delbrück (1900: 98), the place of the genitive forms in PIE "ist dieselbe, wie die des einfachen Adjectivums",[34] hence he supposes a situation where the order was GN & AN (see also Elerick 1991). Preposing genitives are commonly attested in Sanskrit, Ancient Greek, Hittite, and in the Sabellic languages as well: Rosenkranz's (1933) data confirm that here the normal order was GN, not only with pronominal genitives, but also with other possessors, as can be seen in (11):

(11) a. *erar* *nomneper*
of.her:GEN.SG noun.for:ACC.SG
'for her name'
(*Tabulae Iguvinae*, VI a, 23)

b. *senateís* *tanginúd* = Lat. *senatuos sententiad* (SCB[35] 17)
senate:GEN.SG judgement:ABL.SG
'by the judgement of the senate'
(*Cippus Abellanus*[36] 8)

The second example comes from the so-called *Cippus Abellanus*, whose text presents a single example of the reverse order, as illustrated in (12):

(12) *sakaraklúm Herekleís* 'the temple of Hercules'
(*Cippus Abellanus*, 11)

Recent studies agree with Rosenkranz's statement that postnominal genitives are limited to specific usages or contexts, and can be accounted for either by emphasis (in local indications, dedications, or religious references), or by the length of the complement, according to Behagel's law (Bauer 1995: 52–53).

6.2 Looking at Early Latin data, we find a similar situation: postposing genitives are rare, and usually found in topographic indications or names of gods,[37] as in (13):

(13) a. *apud* *aedem* *Duelonai*
at temple:ACC.SG Bellone:GEN.SG
'at the temple of Bellone'
(CIL I^2 581, 1, 2)

---

[34] "The same as the simple adjective".

[35] Although the *Senatus Consultum de Bacchanalibus* dates to 186 B.C.E., its language is extremely conservative.

[36] This limestone tablet inscribed on both sides was found in 1745 at Avella (near Avellino), it was in use as a doorstep and is believed to have been brought from the ancient site of Abella. It is dated to the second century B.C.E., approximately.

[37] "It would seem that in the official religious language it was a customary act of pity to throw the name of god into relief" (Adams 1976: 76).

b. *aedem* *Castoris* *Pollucis*
temple:ACC.SG Castor:GEN.SG Pollux:GEN.SG
'the temple of Castor and Pollux'

(CIL 1, 1506)

Statistical data from Plautus onward show that the GN order no longer was the unmarked option, but despite the clearer preference for NG sequences in later authors, evidence suggests that in Classical Latin multiple factors combine in determining constituent order. Devine and Stephens' fine-grained analyses, however, turn out to be quite frustrating, for it seems that within each category of genitive different words have different distribution and rules.[38] Summing up their findings, they propose different approaches "for reducing this anarchy to a coherent system" (Devine and Stephens 2006: 380), and with regard to the interface between pragmatics and syntax, they observe:

> The crucial factor seems to be conceptual individuation, the degree to which the genitive plus head combination expresses a single recognizable concept, as opposed to an ad hoc combination of two independent concepts.
>
> (Devine and Stephens 2006: 388)

In fact, in some of the oldest GN genitive constructions the tight relationship between the genitive noun and the head noun specifically defines a single entity. In Early Latin such structures are typically found in the following contexts: with kinship terms such as *filius*,[39] for instance in the standard onomastic formula in (14):

(14) *P. Rutilius M. f.* 'Publius Rutilius, son of Marcus'

In the 'emphatic' expressions in (15), which shift toward the reverse order only from Christian times onward (e.g. *vanitas vanitatum* 'vanity of vanities'):

(15) a. *reliquiarum reliquias* 'the remnants of the remnants'

(Pl. *Curc*. 388)

b. *divum deo* '(to) the god of gods'

(*Carm. Sal*. in Varr. *LL* 7, 27)

In the old formulaic and juridical phrases illustrated in (16):

(16) a. *deorum* *beneficio*
god:GEN.PL help:ABL.SG
'by the help of the gods'

b. *vitae* *necis=que* *potestas*
life:GEN.SG death=and:GEN.SG power:NOM.SG
'power of life and death'

---

[38] With regard to the correlation between 'movement' and pragmatics, for instance, "many complements that apparently do not move but stay to the right of the head are also focused, and some complements that do move to the left of the head are not focused" (Devine and Stephens 2006: 380).

[39] Where "the nucleus of information is almost always the identity of the complement rather than the nature of the relationship" (Devine and Stephens 2006: 355).

And in fixed structures like those in (17), which were destined to become compounds because the genitive construction is more likely to be interpreted as a single concept, a structural unity.

(17) *iuris consultus* 'iurisconsult'
manus missio 'manumission'
plebis scitum 'plebiscite'
aquae ductus 'aqueduct'

On the other hand, in early documents the reverse NG pattern occurs not only with the names of gods, but also in a set of noun phrases indicating public functions, as in the following examples:

(18) *pater familias* 'father of the family'
praefectus urbis 'prefect of the city'
magister equitum 'chief of the cavalry'
magister navis 'captain of a ship'
tribunus plebis 'tribune of the people'
tribunus militum 'commander of the army'

For Rosenkranz (1933: 139), in such structures "liegt ein latenter Gegensatz",[40] hence postposition is due to the stress on the element that expresses the opposition. In other words, when less predictable associations of two distinct entities and/or accidental properties are introduced, postnominal genitives are preferred.

But maybe competing patterns are also diachronically motivated. As will be explained in the next paragraphs, two main correlations have been propounded with regard to the functions of the genitive in Latin: a correlation with the verbal phrase and a correlation with the adjective.

6.3 In Benveniste's opinion, the peculiar and original function of genitive constructions can be seen in nominalizations that transform a verbal phrase into a noun phrase: "[l]a fonction du génitif se définit comme résultant d'une transposition d'un syntagme verbal en syntagme nominal" (Benveniste 1966: 148).

This development can be observed in the example (19), where the erstwhile verb turns into the head noun and the genitive modifier transposes the object of a corresponding finite clause.

(19) Finite clause, referring O, OV order
aquam ducere 'to canalize water'
Nominalized VP, referring O, GN construction
aquae ductus 'duct of water'
Nominalized VP, non-referring O, NN structure
aquaeductus 'water duct, aqueduct'

[40] 'There is an underlying opposition'.

Since these structures grammaticalize during a period of OV syntax, older genitive constructions and compounds display a GN order. For Benveniste, this basic transposing relation spreads to nominalizations from intransitive verbs as well, as can be seen in (20), and then also extends from nominalized structures to 'true' possessives, thus originating all the genitive uses.

(20) *sol oritur* 'the sun rises' > *solis ortus* 'the rise of the sun'

As will be shown below, however, 'true' possessives seem to follow a different path.

6.4 According to Woodcock's definition:

> The word or words in the genitive define, describe, or classify the thing or person denoted by the noun qualified. The genitive inflection thus turns a noun or a pronoun into a sort of indeclinable adjective, which is sometimes interchangeable with an adjective.
>
> (Woodcock 1959: 50)

The above quotation entails two relevant assumptions. The first is that, as clearly stated also in Kühner and Stegmann's description of the *attributives Satzverhältnis* (1976 vol. 1: 206), the genitive is viewed as one of the forms whereby the attributive relation can be manifested.[41]

In connection with this, the second assumption is concerned with the overlap between adjective and genitive, which is reminiscent of Wackernagel's claim that in Indo-European languages the use of adjectives expressing possession was not only preferred, but even anterior to genitive constructions:

> In den klassischen Sprachen ist hier durchweg das Adjektiv das Primitivere; der Genetiv mehr der jüngern überhaupt analytischen Sprachstufen eigen.
>
> (Wackernagel 1908: 145–146[42])

Numerous examples show that in Latin this usage is still preserved when nouns of gods and relevant individuals, or nouns indicating social and family relationships are involved, and this seems to explain many interesting cases where NG constructions display corresponding adjectival constructions, such as *sacerdos Veneris* (Pl. *Rud*. 433) and *sacerdos Veneria* (Pl. *Rud*. 329 and 644), or *tribunus militum* and *tribunus militaris* (Cic. *Cato* 32).

In my view, the relevant property of both constructions is not to express possession, but to identify an entity by contrasting it with others and/or by specifying its relationship with another entity, thus combining two distinct concepts. Accordingly, the adjectival expression *amica erilis* 'the master's lover' (Pl. *Mil*. 114 and 121) is consistently paralleled by the genitival construction *amica eri* (Pl. *Mil*. 105 and 262). The reverse order is otherwise chosen when inherent properties pertaining to object description and concept formation are described: by way of illustration, consider the

---

[41] See also the discussion in Touratier 1991.

[42] 'In classical languages is here the adjective the absolute primitive; the genitive pertains more to recent, and predominantly analytic language stages'.

invariable phrase *erilis filius*, which is significantly lacking a corresponding genitival construction[43] and precisely means 'the young master' rather than 'the master's son', thus evoking a single concept (cf. *nostro erili filio* 'to our young master' Pl. *Epid*. 164).

All of this is in perfect keeping with what was said in the preceding section about the relationship between position and bondedness of attributes. Following Wackernagel's insight, it is thus reasonable to assume that, when the genitive supersedes the adjective in such constructions, the new structure parallels the older one, and postnominal genitives display the same loose syntactic bond as postnominal adjectives. Moreover, as Givón points out, in languages in which adjectives are morphologically noun-like, the adjectival pattern can be extended to genitival modifiers: "[u]nlike the abstract *x-bar* principle ('harmony'), these analogical extensions are much more concrete and plausible, driven by close morphological and syntactic similarities" (Givón 2001: 258).

6.5 To sum up, the old GN order presupposes structures with a higher level of integration, and pertains to more predictable and conventionalized possessibility (see verb-phrase nominalizations), which also relates to inherent or inalienable possession (see patronymic formulae), and/or to descriptive, kind referring, non-referential genitives (see compounds).[44] The NG genitive order, on the other hand, is associated with less predictable/accessible and more referential possessors, which combine with their head noun equally contributing an additional item of information within structures with object referring function and a relatively low degree of integration.

In Caesar's works, for instance, the nouns *castra* 'camp' and *copiae* 'forces' tend to precede proper name genitives, but normally follow in the phrases *hostium castra* 'enemy camp', or *hostium copiae* 'enemy forces' (Devine and Stephens 2006: 361–368). Besides, the high rate of postnominal genitives with kinship terms such as *uxor* 'wife', *avunculus* 'maternal uncle', and *patruus* 'paternal uncle' (Devine and Stephens 2006: 355–356), is hardly surprising considering that here the nucleus of information is the nature of the relationship (two concepts), and not someone's identity (one concept), as in the patronymic formula. The same criteria (i.e., relationship vs. identity) seem to govern word order choice when the term *filius* combines with *regis*, but later authors can still adopt the GN pattern also to convey the precise meaning 'son of a king'.[45]

---

[43] The phrase occurs 24 times in Plautus and Terence and is also found, still unchanged, in Suetonius and Gellius. Curiously enough, Hofmann and Szantyr (1965: 408–409) claim: "die Nachstellung, zumindest der possessiven Gen., ursprünglich sein dürfte, also *eri filius* wie *filius erilis*", but both expressions are never attested in Latin! A careful check of the huge corpus of texts in CLCLT-6 provides only the following example: *Ut eri sui corrumpat et rem et filium?* 'to be ruining both the estate and the son of his master?'(Pl. *Most*. 27).

[44] See Dryer (2007b: 177–191). For a detailed discussion of descriptive genitives in English, and on the gradience within *s*-genitives and between *s*-genitives and N+N sequences, see Rosenbach 2006.

[45] Here the passages in question: *eques Romanus es: et ego regis filius*. 'So you're a Roman knight! Well, I'm a king's son!' (Petr. *Sat*. 57, 4); *una feminarum in omni aevo Lampido Lacedaemonia*

As times goes by, the NG pattern starts to encroach on the older one, and this seems to explain why, for instance, *solis ortus* coexists with *ortus solis*. But some apparent inconsistencies can also be due to the clash between pragmatic, semantic and syntactic factors, in a situation partially comparable to the use of genitive constructions in English, where recent analyses clearly pinpoint overlapping areas between coexisting patterns and reconstruct evolutionary trends that motivate grammatical variation (Rosenbach 2005, 2006).

7 Comparative Constructions

7.1 With regard to comparative constructions, Latin displayed two alternatives: the case comparison with the order Standard-Adjective and a particle construction with *quam* and the order Adjective Standard.

The former, also called *ablativus comparationis*, is limited to the following contexts: proverbial and stereotyped phrases, as illustrated in (21):

(21) *melle dulcior* 'sweeter than honey'
luce clarior 'brighter than light'
lapide silice stultior 'more senseless than a flint-stone'
opinione melius 'better than expected'

Negative expressions of comparison, especially negative relative sentences and rhetorical questions that imply a negative answer:

(22) a. *nihil* *est* *virtute* *amabilior*
nothing be:PRS.3PS virtue:ABL.SG attractive:CMP
'nothing is more attractive than virtue'
(Cic. *Lael*. 8, 28)

b. *sequamur* [...] *Polybium nostrum,* *quo* *nemo*
follow:SBJV.1PL P.:ACC.SG our:ACC.SG REL.ABL.SG nobody
fuit [...] *diligentior*
be:PRF.3SG accurate:CMP
'let us follow our Polybium, than whom no one was more accurate'
(Cic. *Rep*. 2, 27)

c. *quis* *me* *est* *ditior?*
who:NOM.SG I:ACC.SG be:PRS.3SG rich:CMP
'who is richer than me?'
(Pl. *Aul.* 809)

Contexts wherein numerical structures with *plus*, *minus*, *amplius*, etc. occur:

---

reperitur, quae regis filia, regis uxor, regis mater fuerit 'Lampido, a Lacedæmonian lady, is the only woman that ever was known, to have been the daughter of a king, a wife of a king, and mother of a king' (Plin. *Nat. hist*. 7, 132); *fugitivam regis filiam, Veneris ancillam, nomine Psychen* 'a king's fugitive daughter, the servant of Venus, named Psyches' (Ap. *Met*. 6, 8).

(23) *palus* *non latior* *pedibus* *quinquaginta*
swamp:NOM.SG not broad:CMP foot:ABL.PL fifty
'a swamp not broader than fifty feet'
(Caes. *BG* 7, 19, 1)

7.2 The particle construction, on the other hand, is the far commoner method for expressing comparison in all periods of Latin: "while the ablative is the original method of denoting comparison, yet in Latin, *quam* with the appropriate case is already well established in our earlier extant literature" (Bennett 1914: 292). An example is given in (24):

(24) *ignoratio* *futurorum* *malorum* *utilior*
ignorance:NOM.SG future:GEN.PL evil:GEN.PL useful:CMP
est *quam* *scientia*
be:PRES.3SG than knowledge:NOM.SG
'ignorance of the future evils is better than knowledge (of them)'
(Cic. *Div*. 2, 9)

According to Bauer, the analytic form replaced the case comparison starting from complex contexts featuring more than two elements, and then spreading to almost all instances, featuring nouns, adjectives, or even clauses. The gist of her argument is that this definite syntactic difference favored the right branching structure, but that between the two types of comparison there was no semantic contrast (Bauer 1995: 140–159).

In my view, Benveniste's idea of a difference in meaning motivating constructions that coexist right from the beginning is far more convincing. The examples in (21), he says, find parallels in other Indo-European languages (e.g. Skr. *svādóḥ svādīyaḥ* 'sweeter than sweet', or Gk. *μέλιτος γλυκίων* 'sweeter than honey', etc.), and can be traced back to the original use in equative comparisons.

Thus, the case construction is a comparison of 'referential adequation' where the referent is an unchangeable standard:

> la construction avec cas donne un comparatif de nature organique et de fonction adéquative, impliquant dans le terme comparant une qualité intrinsèque et prêtant à des emplois 'exemplaires'.
> (Benveniste 1948: 141[46])

On the other hand, the particle construction is as well found in other languages, where adversative (Gk. ἤ,) or negative (Skr. *na*, Gmc. *na*) particles are employed to express a disjunctive comparison that opposes two terms by their extrinsic inequality:

> la construction avec particule donne un comparatif de nature mécanique et de fonction disjonctive, servant à contraster deux termes mis en alternative par une inégalité extrinsèque.
> (Benveniste 1948: 141[47])

---

[46] 'The case construction gives a comparative with organic nature and equative function, involving an intrinsic quality in the comparing term and suitable for 'exemplary' uses'.
[47] 'The particle construction gives a comparative with mechanic nature and disjunctive function, used to contrast two terms opposed by their extrinsic inequality'.

Moreover, whenever the comparative presupposes an unchangeable standard, no restrictions and contextual determinations are allowed, otherwise the *quam* alternative is required:

> une norme est fixe et constante, elle exclut toute restriction ou détermination occasionnelle. Aussi a-t-on *quam*, et non l'ablatif, quand le second terme est accompagné d'une détermination.
>
> (Benveniste 1948: 139[48])

This can be seen in the following example:

(25) *nil* *est* *miserius* *quam* *animus*
nothing be:PRES.3SG wretched:CMP than mind:NOM.SG
hominis *conscius*
man:GEN.SG aware:NOM.SG
'nothing is more wretched than the mind of a man with a guilty conscience'
(Pl. *Most*. 544)

7.3 Interestingly, the view that right branching comparison relates to terms of reference that are semantically more definite and specific, is indirectly confirmed by the following examples:

(26) a. *dulcior* *illo /* *melle*, *quod*
sweet:CMP that:ABL.SG honey:ABL.SG REL.ACC.SG
in ceris *Attica* *ponit* *apis*
in wax:ABL.PL Attic:NOM.SG put:PRES.3SG bee:NOM.SG
'sweeter than the honey that the Attic bee stores in the hive'
(Ov. *Trist.* 5, 4, 27)

b. *durior* *et* *ferro*, *quod* *Noricus*
hard.CMP and iron.ABL.SG REL ACC.SG of.Noricum:NOM.SG
excoquit *ignis*
burn:PRES.3PS fire:NOM.SG
'harder than the iron that the fire of Noricum refines'
(Ov. *Met.* 14, 712)

c. *stultior* *stultissumo*
fool:CMP fool:SUP
'fool of fools'
(Pl. *Amph*. 903)

d. *quis* *clarior* *in Graecia* *Themistocle?*
who famous:CMP in Greece:ABL.SG Themistocles:ABL.SG
'who was more famous than Themistocles in Greece?'
(Cic. *Lael*. 42)

---

[48] 'There is a fixed and invariable pattern that rules out any contextual restriction or determination. And also, one finds *quam* instead of the ablative whenever the second term is accompanied by a determiner'.

Here the case construction displays the reverse order Adjective-Standard, like in the quotation from Caesar in (23). For Bauer (1995: 149), "evidence from Plautus on seems to reflect the uncertainties of the Romans, who were no longer sure about the place and the exact nature of this element" that is, of the term of reference. In my opinion, however, the above examples can signal that, in specific contexts, the order Adjective-Standard was available for the case comparison as well.

As Benveniste correctly argues, the conflict between the two comparative types starts on the borderline, where demonstrative or personal pronouns are involved: here the matter is not adequation versus disjunction, but the contrast between pragmatic and semantic factors, for the term of reference is accessible, but at the same time also definite, specific and animate. And this can be seen from the example in (27).

(27) *nemost* *miserior* *me*
nobody.be:PRES.3SG unfortunate:CMP I:ACC
'nobody is more unfortunate than me'

(Ter. *Heaut*. 263)

These features are indeed more frequent with the *quam*-alternative, which by its nature is employed in wider and more variegated domains and gradually generalizes at the expense of the case comparison.

8 Conclusions

Of course, the analysis presented here is far from being complete, but the data seem to converge on some plausible processing factors influencing grammatical variation and change. Let me briefly recapitulate the details.

As clearly shown in numerous studies, statistics confirm that right branching structures emerge early in Latin. As we have endeavored to explain, instead of assuming a precocious shift to an SVO type (Bauer 1995), possibly limited to spoken language and inconsistently surfacing in written texts (Adams 1976), it is preferable to justify these phenomena as instances of pragmatically-motivated flexibility working on diachronically-mediated patterns. To make the point in a better way, it may be said that structures developing from different sources are assigned different functions according to multiple factors.

Under this view, the overall fluidity of the syntactic cohesion within the Latin nominal group (Lehmann 1991), reflects the relative autonomy of the dependent in dependent-marking languages (Nichols 1986), and is crucially related to word-order flexibility in two ways. On a synchronic level it allows the use of either tighter or looser syntactic bonds depending on the functional, informational and semantic properties of the attributes. In a diachronic perspective, on the other hand, the use of patterns with inconsistent word order paves the way for the shift from the synthetic type to the analytical one.

This phenomenon seems to play a crucial role at the interface between the rise of a fixed word order and the erosion of morphology. In fact, the appearance of analytical structures is not strictly, or directly correlated with the loss of cases, which in itself hardly represents the triggering mechanism of word order change (Sasse 1977: 126, Bauer 1995: 5–10, Magni 2000: 7–9).

For instance, adpositions are often said to be an alternative strategy to express the multifaceted functions proper to overlapping or disappearing cases (Baldi 1979, Cuzzolin, Putzu and Ramat 2006). As we will see in a moment, however, analytic structures were indeed the only viable strategy.

As a matter of fact, "affixal morphology can move in only one direction, from dependent to head" (Nichols 1986: 104), and this precisely happens when forms of adverbial origin become bound preverbs. Increased dependent-marking that reverses the natural process of headward migration via cliticization, boundary reduction and boundary shifting, on the other hand, can occur only under restricted circumstances: in particular, the change of adpositions into affixes is favored by consistent, polar word-order types.[49] This is because if, at the level of NP, the word order is inconsistent, both sequences PREP + MOD + NOUN and NOUN + MOD + POSP prevent the reanalysis of word boundaries as morpheme boundaries (Nichols 1986: 88–89), thus blocking the renewal and recreation of synthetic structures.

In this sense, the use of both prenominal and postnominal attributes can be held to be responsible for the emergence of analytical structures in Latin.

In fact, the agglutination of postpositions is quite restricted even in Umbrian, where it seems independent of case loss,[50] and plausibly conditioned by prosody, according to the recurrent use of structures like *destruco persi* (see (7b), many other examples in Nocentini 1992: 231–233). These sequences are also peculiar of Latin (e.g. *summa cum laude*), where the enclitic status of the adposition can be supposed as well (Nocentini 1992: 236), and where the adding to the morphological case could also signal the tentative change from double-marking to dependent-marking of the tighter syntactic bond within AN patterns.

In general, that postpositions occur in fixed expressions or with personal and relative pronouns, seems to strengthen the point that prehead position pertains to more predictable/accessible information, and mainly to 'light' modifiers.

Postposed coordinators, GN structures, and case comparatives further substantiate the idea that old OV patterns are mostly related to referential predictability/accessibility. On the other hand, the insight that alternative patterns may arise from condensation of erstwhile loose constructions that grammaticalize independently of the OV/VO dichotomy is corroborated by postpositions coming from relational nouns, preposed coordinators, postnominal attributes in NA and NG structures, particle comparatives.

---

[49] Not surprisingly, for instance, among the western languages of the Uralic family, the process of adding to the inherited case inventory by accretion of postpositions is favored by a rigid modifier-head order (Nichols 1986: 89).

[50] Nocentini (1992: 109) observes that, on the contrary, in cases like *asamař* (quoted in 2.2) the agglutination of the postposition favors the maintenance of the accusative ending.

All of this leads us to the conclusion that the alleged 'anarchy' of Classical Latin essentially mirrors the multiplicity of mechanisms interfering in the use of alternating patterns, whose reanalysis is focused on semantic parameters such as definiteness, referentiality and animacy,[51] and on syntactic factors such as weight and complexity. Natural coordination, for instance, occurs in phrases, rather than in clauses. Qualifying adjectives are simple forms; classifying adjectives are instead derivational, more complex and 'nouny'. With prenominal genitives the possessors often represent types rather than tokens of the referent, whereas postnominal genitives are more definite, referential, and complex. And the same holds true for the standards in *quam*-comparatives, which likewise tend to display a more complex syntax.

The gradual re-rigidification of the new order attested in later stages (and also in more variegated text types), seems in its turn to reflect the increasing frequency of patterns with more informative structure and hence the preferred postnominal use of more complex attributes, which naturally have more weight[52] and are more suited to a low degree of bondedness.

In this respect, Lehmann also argues that postnominal position is universally the preferred position of relative clauses, because they contain a predication based on the verb, and are thus more suited to discriminative than to characterizing functions. Therefore, "the Latin relative clause is, from a typological point of view, quite an unremarkable kind of relative clause" (Lehmann 1991: 223) and, *pace* Bauer, cannot be assumed as the turning point of "a linear and irreversible development from left to right branching" (Bauer 1995: 167).

In general, a similar train of reasoning invites caution, for more recent research indicates that the conflation of individual correlations into holistic models[53] leads to unsupported synchronic and diachronic predictions, and the gathering of more extensive data on the languages of the world also casts doubts on chains of correlations.

The evolution at issue is actually all but linear, and probably lacking a *primum movens*. Basically, the overall picture of Latin where protracted and overlapping transitional stages generate internal and typological inconsistencies is due to the coexistence and mixing of different patterns, and to the different rate at which old structures disappear, sometimes leaving "[a] recalcitrant residue of diachronically motivated exceptions [that] will continue to haunt even the most powerful synchronic generalizations we can find" (Givón 2001: 234).

---

[51] The peculiar role of semantic parameters in determining the increasing frequency of postverbal objects, as well as its interface with the erosion of morphology, has been discussed in Magni 2000.

[52] As Givón puts it: "the more disruptive, surprising, discontinuous or hard to process a topic is, the more coding material must be assigned to it" (Givón 1983: 18).

[53] To be honest, a similar approach has been considered in Magni (2000: 27–29) too, where adjectival and genitival patterns were attributed a relevant role, which is partly consistent with the present considerations about the behavior of attributes.

References

Adams, James N. 1976. A typological approach to Latin word order. *Indogermanische Forschungen* 81: 70–99.

Baldi, Philip. 1979. Typology and the Indo-European prepositions. *Indogermanische Forschungen* 84: 49–61.

Bauer, Brigitte L.M. 1995. *The emergence and development of SVO patterning in Latin and French*. Oxford: Oxford University Press.

Bennett, Charles E. 1914. *Syntax of early Latin*. Vol 2. *The cases*. Boston: Allyn and Bacon.

Benveniste, Emile. 1948. *Noms d'agent et noms d'action en indo-européen*. Paris: Maisonneuve.

Benveniste, Emile. 1966. *Problèmes de linguistique générale*. Paris: Editions Gallimard.

Bhat, D.N. Shankara. 1994. *The adjectival category*. Amsterdam/Philadelphia: John Benjamins.

Bolkenstein, Alide Machtelt. 2001. Random scrambling? Constraints on discontinuity in Latin noun phrases. In *De lingua Latina novae quaestiones. Actes du X^{è} Colloque International de Linguistique Latine*, (*Paris-Sèvres, 19–23 avril 1999*), ed. Claude Moussy, 245–258. Louvain/Paris: Peeters.

Buck, Carl Darling. 1904. *A grammar of Oscan and Umbrian*. Boston: Ginn and Co. Publishers.

Bybee, Joan L. 2006. Language change and universals. In *Linguistic universals*, eds. Ricardo Mairal, and Juana Gil, 179–194. Cambridge: Cambridge University Press.

Carvalho, P. de. 1991. Structure morpho-sémantique de l'adjectif latin ou: la presence de l'absent. In *New studies in Latin linguistics*, ed. Robert Coleman, 251–267. Amsterdam/Philadelphia: John Benjamins.

CLCLT-6. 2005. *Library of Latin texts*, Centre 'Traditio Litterarum Occidentalium', ed. Paul Tombeur. Turnhout: Brepols. http://www.brepolis.net/info/info_clt_en.html

Cuzzolin, Pierluigi, Ignazio Putzu, and Paolo Ramat. 2006. The Indo-European adverb. *Indogermanische Forschungen* 111: 1–38.

Delbrück, Bertold. 1900. *Vergleichende Syntax der indogermanischen Sprachen*. 3 Teil. Strassburg: Trübner.

Del Vecchio, Tommaso. 1989. Observations sur l'ordre des mots dans la latin archaïque. In *Subordination and other topics in Latin. Proceedings of the third colloquium on Latin linguistics* (*Bologna 1–5 April 1985*), ed. Gualtiero Calboli, 541–555. Amsterdam/Philadelphia: John Benjamins.

Devine, Andrew M. and Laurence D. Stephens. 2006. *Latin word order. Structured meaning and information*. Oxford: Oxford University Press.

Dryer, Matthew S. 1988. Object-verb order and adjective-noun order: Dispelling a myth. *Lingua* 74: 185–217.

Dryer, Matthew S. 1992. The Greenbergian word order correlations. *Language* 68: 81–138.

Dryer, Matthew S. 2007. Word order. In *Language typology and syntactic description. Second Edition, Vol. I: Clause structure*, ed. Timothy Shopen, 61–131. Cambridge: Cambridge University Press.

Dryer, Matthew S. 2007. Noun Phrase Structure. In *Language typology and syntactic description. Second Edition, Vol. II: Complex constructions*, ed. Timothy Shopen, 151–205. Cambridge: Cambridge University Press.

Dryer, Matthew S. 2008a. Order of subject, object and verb. In *The world atlas of language structures online*, eds. Martin Haspelmath, Matthew S. Dryer, David, Gil, and Bernard Comrie, ch. 81. Munich: Max Planck Digital Library. http://wals.info/feature/81. Accessed 20 Apr 2008.

Dryer, Matthew S. 2008b. Relationship between the order of object and verb and the order of adposition and noun phrase. In *The world atlas of language structures online*, eds. Martin Haspelmath, Matthew S. Dryer, David, Gil, and Bernard Comrie, Ch. 95. Munich: Max Planck Digital Library. http://wals.info/feature/95. Accessed 20 Apr 2008.

Elerick, Charles. 1991. Latin noun/gen./adj. serialization and Language Universals. In *New studies in Latin linguistics*, ed. Robert Coleman, 311–322. Amsterdam/Philadelphia: John Benjamins.

Ernout, Alfred and François Thomas. 1951. *Syntaxe latine*. Paris: Klincksieck.
Foley William A. 1980. Toward a universal typology of the noun phrase. *Studies in language* 4.2: 171–199.
Friedrich, Paul. 1975. *Proto-Indo-European syntax: The order of meaningful elements*. *Journal of Indo-European studies*, Monograph 1. Butte: Montana College of Mineral Science and Technology.
Giusti, Giuliana, and Renato Oniga. 2006. La struttura del sintagma nominale latino. In *Atti della "Giornata di Linguistica Latina"* (*Venezia, 7 maggio 2004*), eds. Renato Oniga and Luigi Zennaro, 71–100. Venezia: Cafoscarina.
Givón, Talmy. 1980. The binding hierarchy and the typology of complements. *Studies in language* 4.3: 333–377.
Givón, Talmy. 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam/Philadelphia: John Benjamins.
Givón, Talmy. 2001. *Syntax: An introduction.* Amsterdam/Philadelphia: John Benjamins.
Gonda, Jan. 1954. The history and original function of the Indo-European particle $^{*}k^{w}e$, especially in Greek and Latin. *Mnemosyne* 4–7: 177–214 (part I), 265–296 (part II).
Greenberg, Joseph H. 1969. Some methods of dynamic comparison in linguistics. In *Substance and structure of language*, ed. Jaan Puhvel, 147–203. Berkeley: University of California Press.
Hofmann, Johann B., and Anton Szantyr. 1965. *Lateinische syntax und stilistik*. München: Beck.
Humbert, Jean. 1960. *Syntaxe grecque* (3^{e} *edition, revue et augmentée*) Paris: Klincksieck.
Kühner, Raphael, and Carl Stegmann. 1976. *Ausführliche Grammatik der lateinischen Sprache. Zweiter Teil*. *Satzlehre*. 2vls, 5th ed., repr. 1992. Darmstadt: Wissenschaftliche Buchgesellschaft.
Lehmann, Christian. 1979. Zur Typologie des Lateinischen. *Glotta* 57: 237–253.
Lehmann, Christian. 1991. The Latin nominal group in typological perspective. In *New studies in Latin linguistics*, ed. Robert Coleman, 203–232. Amsterdam/Philadelphia: John Benjamins.
Lehmann, Winfred P. 1972. Contemporary linguistics and Indo-European studies. *PMLA* 87: 976–993.
Lehmann, Winfred P. 1974. *Proto-Indoeuropean syntax*. Austin: University of Texas Press.
Linde, Paul. 1923. Die Stellung des Verbs in der lateinischen Prosa. *Glotta* 12: 153–178.
Magni, Elisabetta. 2000. L'ordine delle parole nel latino pompeiano: sulle tracce di una deriva. *Archivio glottologico italiano* 85: 1–35.
Magni, Elisabetta. forth. Conservazione e innovazione nella morfologia derivazionale dell'italiano: analisi sincronica e diacronica del suffisso *-aio*. In *Prospettive nello studio del lessico italiano. Atti del IX congresso internazionale della società di linguistica e filologia italiana* (*Firenze 14–17 giugno 2006*), eds. Emanuela Cresti, and Maurizio Moneglia.
Marouzeau, Jules. 1922. *L'ordre des mots dans la phrase latine*, *Tome I: Les groupes nominaux*. Paris: Champion.
Marouzeau, Jules. 1949. *L'ordre des mots dans la phrase latine*, *Tome III: Les articulations de l'enoncé*. Paris: Les Belles Lettres.
Marouzeau, Jules, 1953. *L'ordre des mots en latin*. Volume complémentarie. Paris: Les Belles Lettres.
Miller, D. Gary. 1975. Indo-European: VSO, SOV, SVO, or all three? *Lingua* 37: 31–52.
Nichols, Joanna. 1986. Head-marking and dependent-marking grammar. *Language* 62: 56–119.
Nocentini, Alberto. 1992. Preposizioni e posposizioni in oscoumbro. *Archivio glottologico italiano* 77: 196–242.
Nocentini, A. 2005. Du latin aux langue romanes: la contribution de la typologie. In *Latin et langues romanes. Etudes de linguistique offertes à József Herman à l'occasion de son* $80^{ème}$ *anniversaire*, eds. Sándor Kiss, Luca Mondin, and Giampaolo Salvi, 411–418. Tübingen: Niemeyer.
Panhuis, Dirk. 1984. Is Latin an SOV language? *Indogermanische Forschungen* 89: 140–159.
Rix, Helmut. 2003. Ausgliederung und Aufgliederung der italischen Sprachen. In *Languages in prehistoric Europe*, eds. A. Bammesberger and Th. Vennemann, 147–172. Heidelberg: Winter.

Rosenbach, Annette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language* 81.3: 613–644.
Rosenbach, Annette. 2006. Descriptive genitives in English: a case study on constructional gradience. *English language and Linguistics* 10.1: 77–118.
Rosenkranz, Bernhard. 1933. Die Stellung des attributiven Genetivs im Italischen. *Indogermanische Forschungen* 51: 131–139.
Sasse, Hans Jürgen. 1977. Gedanken über Wortstellungsveränderung. *Papiere zur Linguistik* 9: 82–142.
Stassen, Leon. 2000. AND-languages and WITH-languages. *Linguistic typology*, 4: 1–54.
Touratier, Christian. 1991. Adjectif et syntagme nominal. In *New studies in Latin linguistics*, ed. R. Coleman, 233–249. Amsterdam: Benjamins.
Vennemann, Theo. 1974. Topics, subjects, and word order: from SXV to SVX via TVX. In *Historical linguistics vol. 1: Syntax, morphology, internal and comparative reconstruction*, eds. John M. Anderson, and Charles Jones, 339–376. Amsterdam: North Holland Publishing Company.
Viti, Carlotta. 2006. *And* in the Early Indo-European languages. *Archivio glottologico italiano* 91.2: 129–165.
Wackernagel, Jacob. 1908. Genetiv und Adjektiv. In *Mélanges de linguistique offerts à M. Ferdinand de Saussure*, 125–152. Paris: Champion.
Wälchli, Bernhard. 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.
Woodcock, Edward C. 1959. *A new latin syntax*. London: Methuen and Co. Ltd.

Typological Universals and Second Language Acquisition

Anna Giacalone Ramat

Abstract In the first part of this paper I deal with some points concerning the relationship between typology and Second Language research that were touched upon by Joseph Greenberg in his 1991 article: the contribution of L2 research to linguistic theorizing, the nature vs. history dichotomy, the relation between second languages and pidginization processes. Next, I point out the relevance of typological universals to accounts of learner behavior. Finally, the issue of tense and aspect marking in second languages is addressed in the light of the theoretical proposal known as the "Primacy of aspect hypothesis". The explanatory power of this hypothesis is discussed by drawing on the notion of prototype, which is in turn argued to be in need of further refinements.

Keywords Greenberg's universals · second language acquisition · tense/aspect marking · semantic properties of predicates

1 Greenberg and Second Language Acquisition[1]

In a book edited by Thomas Huebner and Charles Ferguson in 1991 (*Crosscurrents in Second Language Acquisition and Linguistic Theories*), Greenberg provided some reflections on the relation between typology and universals and second language acquisition (henceforth SLA) research. These pages are not widely known, although they are full of notable remarks and show how open-minded he was toward new fields of research.

A. Giacalone Ramat
University of Pavia, Italy

[1] I met Joseph Greenberg for the first time in Los Angeles at the fourth International Conference on Historical Linguistics (1979). On that occasion he supported the proposal of organizing an ICHL conference in Europe, in particular in Pavia (it was realized as the Seventh ICHL, in 1987). I was very impressed by his extraordinary knowledge of languages and his insightful thoughts.

S. Scalise et al. (eds.), *Universals of Language Today*, 253–272,
DOI: 10.1007/978-1-4020-8825-4_13,

In his 1991 article Greenberg points out that:

> The relationship [between the field of typology/universals and that of second language acquisition] has been an unequal one, with typology/universals, based on the cross linguistic study of fully functioning adult languages, as the donor and source of hypotheses in studies of second language acquisition.
>
> Greenberg 1991:37

Thus, Greenberg recognizes a kind of priority to typology – both chronologically and also in terms of importance. In his view, second language acquisition would be in the same position as a number of other fields, such as first language acquisition, language loss in aphasia studies, and pidgin and creole studies.

But in the same article Greenberg takes issue with the long lasting debate concerning the role of second languages (L2s) in linguistic theorizing: first (and contrary to generative linguistics), he does not consider first languages as a privileged field of study in the analysis of the human linguistic capacity in that both L1 and L2 have specific primary languages as reference points; and secondly, he accepts L2 data as a source for language universals (Greenberg 1991:39, see also Hyltenstam 1986:67):

> Put more generally, for all these fields, one may say that universals apply equally to interlanguage and to primary language.
>
> Greenberg 1991:39

This is not surprising, since Greenberg's universals, as is well known, are not properties of the human mind, but inductive generalizations and allow for all kinds of languages as historical products. Consequently, learner languages also belong to the field of typology and are relevant to the validation of language universals.

Further, Greenberg emphasizes that typology has provided a fundamental guiding principle in language studies, namely that in linguistic oppositions at all levels, one member, the "unmarked" one, is favored over the other, the "marked" one. The concept of markedness is indeed a fundamental concept and an important tool not only in typological research: it has been applied both to first and second language research and to the fields of aphasia and creole studies (Greenberg 1991:38).

It is undeniable that typology has been able to produce theoretical tools in the form of implicational generalizations, which provide hypotheses for the description of learner languages. I will just mention the rich literature on the acquisition of relative clauses based on one of the most popular results of typological research, the *Accessibility Hierarchy* (Keenan and Comrie 1979, etc.) and the debate sparked off by Greenberg's universals among SLA researchers (see, among others, Rutherford 1984, Hyltenstam 1986, etc.)

1.1 The Contribution of SLA Research to Research on the Nature of Language and Human Cognition

The approach advocated by Greenberg concerning the highly desirable interaction of typology and SLA has been developed in more recent SLA studies which have

pointed out that SLA research can not only contribute to a better understanding of how people acquire languages, but also to a better understanding of how the human linguistic capacity is structured and functions (Klein 1991:184). In other words, such studies can actively contribute to research into universals and help to clarify theoretical issues.

Within this perspective, I have argued (in some of my relatively recent work, Giacalone Ramat 1995, 2003a and c) that in learner languages, acquired under "naturalistic" circumstances, form-meaning relations are simpler and relatively more transparent than in fully developed languages (the usual object of study of typology) and can better illustrate the organizing principles of language systems as well as markedness relations. Learner languages would represent a set of less marked options. In the second part of this paper I will try to illustrate this issue by means of examples taken from the domain of tense and aspect.

The claim that learner varieties belong to the field of typology has been put forward in a number of recent works, including Hammarberg and Koptjevskaja-Tamm's (2003) investigation into the acquisition of adnominal possession in Swedish, and Bernini's (2003) work on the implicational organization of indefinite pronoun functions in learner Italian, which conforms to the parameters postulated by typologists (Haspelmath 1997). One more example is provided by work on clause combining strategies in learner Italian (Giacalone Ramat 1999, 2003b), which reflect the functional principles governing the correlation of semantic and syntactic dimensions of complementation (Givón's 1990 "Binding Hierarchy"). Researchers working from the functional perspective of the "Basic Variety" have also emphasized the contribution of second language studies to the understanding of the nature and functioning of language (Klein and Perdue 1992, 1997 and related work). According to the Basic Variety approach, learner languages are characterized by a limited set of structural, semantic and pragmatic principles. A case in point is the "controller first" principle: the more agentive referent appears first, usually before the verb (Klein and Perdue 1997). This is in keeping with typological investigations which have introduced the notion of control strength (Comrie 1989: 58ff). The early stages of learner languages do not show modifications of this preferred order in control parameters: passive constructions occur somewhat later, and object topicalization, cleft sentences and other focalizing devices are rare (see Chini 2002 for Italian). We may conclude that the effect of control hierarchies is reflected more directly in learner language preferences than in fully-fledged languages (Giacalone Ramat 2003a: 5).

Such results suggest that SLA research is an especially promising field of research for theoretical linguistics rather than (only) a rich field for all kinds of case studies, as pointed out by Ferguson (1991:434).

1.2 Nature vs. History

Returning to more Greenbergian considerations, I would like to illustrate one more point touched on by Greenberg in his article, namely the relation between two

fundamental dimensions which have often been contrasted in linguistic theorizing. Greenberg calls these two dimensions "nature" and "history" and establishes a dichotomy between them. The "state of nature" in Rousseau's terms – he observes – is "a notion which is old in Western thinking" and is reflected in the notion of "naturalness" in linguistic theories, such as Stampe's natural phonology, and also in Bickerton's view of creoles as expressing the essential nature of human languages because they, as it were, have no history (Greenberg 1991: 39–40). "All of these views are in varying degrees ahistorical" – Greenberg claims – but languages, including interlanguages, are historical products. Greenberg recognizes in these contrasting views a critical point for linguistic research both in SLA and in creole linguistics and suggests that nature and history should be taken as two complementary aspects in language evolution.

In SLA, the dichotomy nature vs. history has often been interpreted as an opposition between universal grammar and transfer in a broad sense, that is in the sense of the effect of the varying social conditions and individual experiences under which second language acquisition takes place. In the field of pidgin/creole studies a similar debate has also flourished in that universalist vs. substratist hypotheses have been contrasted (Muysken and Smith 1986, Romaine 1988, McWhorter 1998, 2001, Holm 2000). Greenberg's claim that the two approaches are not incompatible draws on two important tenets of functional typology, namely that universals are relational and diachronic (recall that one of Greenberg's merits was the "dynamicization of typology") rather than a fixed and static set of principles (1991:41).

As far as the relation between creole and second language acquisition research is concerned, Greenberg's position reflects the debate that flourished in the 1980s and in the 1990s, with changing directions, as to the relevance of one field to the other. To summarize some key points briefly, second language acquisition in conditions of restricted input was considered by some researchers a case of pidginization (Schumann 1978, Andersen 1983). Some decades of research have provided a more accurate body of knowledge about the complex sociolinguistic circumstances in which pidgin and creole languages come into being and have also given us a huge number of studies on the features which are, at least partially, shared by pidgin/creoles and learner varieties (cf. Romaine 1988 for detailed discussion). Recently, Kouwenberg and Patrick, resuming the role of SLA in pidginization and creolization, have recognized that at present:

> the general thrust is to reverse the earlier direction of disciplinary influence by applying knowledge from SLA studies to refine our understanding of the processes at work in natural situations of language contact, creation, and acquisition
>
> Kouwenberg and Patrick 2003:180

But the relevance of SLA processes seems to be limited to early stages of contact, with a cut-off point after which creoles undergo rapid internal changes and turn into autonomous language systems, while SL varieties are strongly oriented toward target language norms.

To be sure, Greenberg's position appears quite in line with recent perspectives which try to reach some sort of compromise between universal and substrate explanations, recognizing that both approaches can contribute to explaining the complex phenomenon of creole genesis. In particular, the field of creole studies would profit greatly from results in neighboring subdisciplines, like acquisition, language contact studies, code-switching and bilingual speech behavior studies (Holm 2000, Arends, Muysken and Smith 1994: 329f).

2 Universals in Typological Research: Exploring Linguistic Diversity

So far I have discussed some points concerning the relation between universals and SLA as stated by Greenberg. Functional typology has basically followed Greenberg's teaching in doing linguistics. For typologists, the path to linguistic universals goes through cross-linguistic generalizations: a typologist uses an inductive method of analysis by constructing a sample of the world's languages and seeking language universals via cross-linguistic generalizations (Croft 2001:7). Thus, within the functional view, universals of human language are not necessarily an innate genetic endowment.[2]

The empiricist method used by typologists to search for universals is essentially shared by SLA studies of the functionally oriented type. The description of linguistic phenomena revealed by acquisitional processes such as relative clauses or tense-aspect marking is rooted in a comparative perspective and a universalistic approach. Typological generalizations account for learner behavior and make predictions about developmental patterns as well as about the relative time of acquisition. The principles that motivate the patterning of grammatical structures and their changes over time are extra-linguistic, semantic and discourse principles (they are basically the same motivating factors as those found in grammaticalization studies). This type of explanation again corroborates the link between SLA and the functional-typological approach. It has been suggested that one of the most distinctive features of functional typology is "the study of all types of linguistic variation – cross-linguistic (synchronic typology), intralinguistic (sociolinguistics and language acquisition) and diachronic (diachronic typology and historical linguistics) in a unifying perspective" (Croft 2003:289).

---

[2] The term 'Universal Grammar' is associated with generative grammarians' claim that certain aspects of syntactic structure are not only universal, but innate. I will use the term universal grammar in a broader sense, without assuming that it is innate (a similar position is found in Croft 2001:9):

> In the broad sense, Universal Grammar consists of those aspects and properties of grammatical structure which form the make-up of the grammars of all human languages
> (Croft 2001:9)

To be sure, Greenberg presented his universals as a set of syntactic constructions governed by some deeper typological pattern like "harmony" (Greenberg 1966), but was not really interested in finding extra-linguistic explanations for the observed generalizations. This attitude sets Greenberg's approach apart from subsequent developments in the field of functional typology, which, though strongly inspired by Greenberg, has drawn attention to universal principles based on human cognition and human communication, which may offer an explanation for many linguistic phenomena. The prevailing opinion of linguists working within the framework of cognitive linguistics or usage-based linguistics is that language structure emerges from language use (Tomasello 2003 for L1, Croft 2003, 2001, Givón 1995, among others). This universalistic approach based on cognition and communication makes many claims and predictions that are easily testable on second language acquisition data. Acquisition can be understood as a process sensitive to individual factors and diverse social conditions, but oriented and guided by universal principles based on human cognition and communication.[3]

Research on Italian as an L2 has been inspired by functional typological studies (Giacalone Ramat 2003c for a general overview). In the following I will try to assess the state of our understanding of L2 acquisition of tense/aspect distinctions. But I will first attempt to clarify some problems concerning morphological coding.

3 The Growth of Morphology[4]

The emergence of morphology in learner languages displays a number of interesting properties and raises a number of issues about the nature of cognitive representations in the learner's mind as well as the role that can be attributed to typological

[3] Among the so-called external factors, frequency and context have been shown to play an important role in the acquisition process. In SLA, a learner is exposed to utterances in context. Consequently, one may say that a significant part of the meaning of the utterance is available from the context of use (this is also the case in child language). And, as is argued by Greenberg, frequency is one parameter for markedness. Consequently, frequency of occurrence appears as a primary factor both in synchronic language states, in language change and in acquisition. Among the well-attested markedness patterns (Croft 1990:64ff), the singular, which has usually less structural marking than the plural, is statistically more frequent than the plural. But only frequent forms can preserve irregularity, while non-frequent irregular forms are often lost, as the dynamics of language change shows. Such patterns are reflected in the expression of inflectional categories in learner languages (also in the case of Italian, see below).

[4] The work devoted by J. Greenberg to morphological typology (1954=1960) should be mentioned here, although it was not as influential as his classification of word order types. Greenberg used a quantitative approach based on indexes of different types: the number of morphemes per word, the types of morpheme (inflectional vs. derivational), and affix position (Giacalone Ramat 1994). Greenberg's aim was to characterize the concept of linguistic type showing that languages need not be classified into discrete types, because the number of morphemes per word is a continuous value (Croft 1990:41). Greenberg's morphological typology was a remarkable enterprise, which deserves to be applied to further data. Some research in the direction of verifying the distributional taxonomies discovered by him has recently been carried out by Cuzzolion, Putzu and Ramat (2006).

universals. Indeed, one of the most intriguing phenomena in learner varieties is the lack of inflectional morphology, which is a crucial component for many languages, albeit, admittedly, to varying degrees. Interestingly, early learner varieties rely on lexical items and basic syntactic constructions[5]. It is well known that the distribution of morphological inflections in languages is extremely varied: there are languages with rich morphology (Slavic languages, German) and languages which are weakly inflecting, such as Romance languages, which have number and gender distinctions in noun morphology, but no case inflections (i.e. word-internal morphology). There are also languages which lack morphology, such as Vietnamese and, to some extent, Chinese. This picture has suggested the conclusion that:

> Morphology is not necessary. There are languages that do without it, and languages with morphology vary quite remarkably in their morphological structure and complexity.
>
> Mark Aronoff, *Morphology by Itself*, 1994:165

By contrast, in learner languages morphology is always absent at earlier stages, and it only develops subsequently as a consequence of the contact with the input, for those languages which have morphology, of course. Thus, early learner varieties represent the zero-degree of morphology, irrespective of whether the target language is rich in morphology or not. This means that learners of morphologically rich languages are exposed from the beginning to morphology, but are not able to handle it. Children are said to be better than adults at learning morphology, although sometimes they overgeneralize grammatical morphemes (adults may do the same, incidentally). Since to learn morphology means to discover a network of form-function relations, the challenge both for SLA researchers and for theoretical linguists and typologists is to find out whether the developmental path or, rather, the possible developmental paths exhibit preferences and regularities reflecting universal principles, such as markedness principles.

Actually, empirical investigations, like those cited above concerning Italian as an L2 (Giacalone Ramat 2003a and c, and references therein), have mostly confirmed the expected markedness patterns. Learners acquire less marked and more frequent forms earlier, in accordance with generalizations based on implicational markedness scales. As an illustration, I will discuss some phenomena related to the category of number. Regardless of details, plural inflections may be absent in the early stages of acquisition, but then learners gradually acquire them without any particular difficulty. However, some errors are found which are, as usual, theoretically important:

(1) plural inflections are frequently missing when quantifiers and numerals are present, as shown in examples (1) and (2), suggesting that, in such contexts, plural marking on nouns is felt to be redundant. This case could be explained as

---

[5] The reasons for the absence of morphology in learner varieties may be varied. Most importantly, morphology is complex and requires significant memory costs. Secondly, it is partially arbitrary (see e.g. the rules for gender assignment in languages such as German or Italian). Thirdly, inflections are often not very salient in the speech stream and are polyfunctional in many cases. Finally, there is compensation for the lack of morphology at the syntactic level (see the summary in Tomasello 2003:240 on child language acquisition).

a classic case of neutralization of number distinctions in certain contexts such as those involving numerals or other quantity words, as is found in Lezgian and Turkish (WALS 2005: 143, Ch. 34 by Haspelmath), or as an overextension of the less marked option. In Italian, both the singular and the plural are marked for number via an overt inflection; however, the singular is typologically unmarked on cross-linguistic grounds (Croft 2003:89).

(2) as for example (3), one must keep in mind that the Italian plural morphemes combine the features of number and gender. In the case at hand, the number of morphological distinctions is reduced, since the masculine morpheme *-i* is overgeneralized: *i personi* = *le persone*, in accordance with the typological markedness pattern for gender: masculine< feminine< neuter (Croft 2003:156).

(1) alcune gita
some-FEM.PL trip-SG Banca dati, AN (L1 German)

(2) tre fratelo Banca dati, CH (L1 Chinese)
three brother-SG

(3) i personi andicappati
the person-PL handicaped-PL
Banca dati, AN (L1 German)

Example (2) is possibly influenced by the first language, which does not allow plural markers after numerals. Thus, from such examples, it follows that L2 learners' developing grammars are congruent with markedness principles, though at the same time obeying other factors, such as token and type frequency, or first language experience which all together enable learners to make their way through the morphology of the target language.[6] Language acquisition studies offer hypotheses on the development of markedness values for grammatical categories.

To learn to express a category in a second language does not simply mean to imitate the input and to relate inflections to some functions, but it may imply what Slobin (1996) has called "thinking for speaking", namely the necessity to conceptualize a category at some pregrammatical level, if the specific language which is being learnt requires obligatory grammatical marking for that category. An example could be the learning of the linguistic distinctions for gender in Italian by Chinese speaking learners, whose first language lacks grammatical gender (Giacalone Ramat 2003d).

It should be made clear that the term "category" is used here in a traditional, neutral way. Currently, there are no widely accepted definitions of linguistic categories,

[6] A different but to my eyes not irreconcilable approach to markedness in morphology is defended by Dressler and researchers working in the Natural Morphology paradigm: morphology is a key that opens the door to establishing functional categories such as Tense and Aspect (Voeikova and Dressler 2002). The emergence of paradigms in children occurs after a pre-morphological stage.

although the issue has often been debated in linguistic theories, both functional and formal. Croft (2001) denies the status of Universal Grammar categories to parts of speech such as nouns and verbs. Be that as it may, to assess the status of grammatical categories is not relevant to our present concerns. I will use the descriptive model proposed by Ramat (1999=2005), according to which a category is viewed as a bundle of features and values.

As I have already pointed out, second language data suggest that features, such as "gender" and "number", and values, such as "masculine" and "singular", are acquired gradually and are also compatible with the hypothesis that the corresponding categories are built up gradually (see also Section 4 on Actionality).

It has been noted in several studies on L2 Italian (Bernini 1990:86, 2003, Giacalone Ramat 1992) that in early learner productions words occur to which it is sometimes difficult to assign a lexical category, such as "verb" or "noun": the category is underspecified. The learner does not seem to realize which properties of the input (like inflections or argument structure) are relevant:

(3) allora mia filia ha telefonata di Germania + + oh mamma come come stai?

then my daughter has called from Germany, mom how how are you?

in Germania hanno detto in Italia tutto pieno di nevica

in Germany have-PRES.3PL said in Italy all full of snow-PRES.3SG

"then my daughter called me from Germany and said: mom, how are you? In Germany they said that in Italy everything is full of *snow"

Banca dati, FR (L1 German) 06

(4) lui vuole riposo
he want-PRES.3SG "rest" (?)

CH, L1 Chinese (Giacalone Ramat 2003d:21)

(5) diploma di domande e risponde
certificate of questions and "answer"-PRES.3SG (?)

MK, L1 Tigrinya (Bernini 1990:87)

The appropriate morphology of the target language is missing: *nevica* in (3) exhibits verbal morphology, but a noun is intended here as shown by the syntactic construction.

The development of verbal morphology and argument frames for predicates also appears to be a gradual process (see Section 4 below).

Finally, to go back to our question of the role of morphology in learner varieties, there is evidence from learners' data that lexical items and basic syntactic constructions do all the work in initial learner varieties. (This situation recalls the case of pidgins and creoles, as noted by McWhorter 1998.) However, morphology gradually emerges in a manner consistent with predictions made by typological universals and learners work out the meaning of the various markings provided by the input on the grounds of general cognitive principles and discourse frequency.

4 The Acquisition and Use of Tense and Aspect Grammatical Structures: The 'Primacy of Aspect' Hypothesis and the Notion of Prototype

As a final topic, I will discuss some issues related to the acquisition of verb morphology from the point of view of the functions conveyed by tense and aspect morphemes. I will argue that findings in these domains confirm what has been said until now on the possible contribution of SLA research to linguistic theories.

My claim is that the patterns of form-meaning associations in the temporal/aspectual domain are acquired by means of both general cognitive (potentially universal) principles and the linguistic context. To support this claim, it will be necessary to investigate how tense and aspect distinctions gradually emerge in learners' productions.

I will first discuss the theoretical proposal known as the "Primacy of aspect hypothesis" and assess its explanatory power, then I will propose a new hypothesis concerning the acquisition of actionality.

According to a widely held opinion, in both L1 and L2 acquisition studies, the acquisition of tense-aspect structures is triggered by actionality distinctions. Despite differences in detail, the main reference in these studies is Vendler's (1967) classification of verbal predicates into four semantic classes, although it has long been recognized that syntagmatic and phrasal properties also have to be taken into account to define the aspectual and semantic properties of predicates (as pointed out, e.g., by Verkuyl 1993 and Pustejovsky 1995). Simplifying somewhat, we may propose the following correlation between the main semantic dimensions involved (see also Bertinetto and Noccetti, 2006).

(8)	telic predicates	perfective aspect	past tense
	atelic predicates	imperfective aspect	present tense

Evidence supporting this correlation comes from a number of investigations into both L1 acquisition and L2 acquisition and also fully developed adult languages (starting in the 1970s with Antinucci and Miller 1976, then Slobin 1985, Tomasello 2003 on L1). In SLA research, the view that tense/aspect morphology correlates with predicate semantics rather than tense has gained wide attention. This position is known as the "Primacy of Aspect Hypothesis" (Andersen 1991, Shirai and Andersen 1995, Andersen and Shirai 1996). A first point to be clarified (perhaps an obvious one) is that "aspect" here does not refer to grammatical aspect, but to the inherent properties of verbal predicates, sometimes (although somewhat misleadingly) referred to as lexical aspect, for which I shall employ the term *Aktionsart* or Actionality (Bertinetto 1986).

The "Aspect Hypothesis" could be in principle valid for all languages which have morphological distinctions of tense and aspect in the verb. However, it must be borne in mind that the implementation of actional and aspectual categories is not uniform in the languages of the world, since languages may not explicitly code all relevant oppositions, but may exhibit cases of ambiguity and neutralization (as is shown by the distribution of the grammatical marking of the perfec-

tive/imperfective distinction in the languages of the world: WALS 2005, Ch. 65 by Dahl and Velupillai). Until now, the Primacy of Aspect Hypothesis for L2 has been tested on a wide variety of languages, including Romance languages, English, Chinese, Japanese (Andersen and Shirai 1996: 543ff for a review, Salaberry and Shirai 2002, Shirai and Nishi 2003, etc.). Corpus data from a wider variety of languages will hopefully be soon available for a more systematic investigation.

In this paper I will focus on Romance languages in which the distinction between imperfective/perfective aspect is marked only in the past, while actionality is mainly specified lexically and requires contextual interpretation. It has been claimed that perfective past marking (Italian *passato prossimo*, French *passé composé*, Spanish Preterit) is initially restricted in learners to telic predicates (achievements and accomplishments in Vendler's terms). By contrast, atelic predicates, states and activities, preferably receive imperfective marking, namely present tense inflections at first, then imperfect inflections. As I said, many empirical studies are consistent with the generalization that the acquisition process is mediated by the semantic properties of lexical verbs (as is confirmed by the recent collection of studies edited by Salaberry and Shirai in 2002). To be sure, language specific factors, such as L1 transfer, input data, etc. may play a role and cause a number of problematic cases in individual developments, but, nonetheless, the driving force of actionality appears to be a robust hypothesis largely supported by empirical evidence.

The explanation for the observed tendencies is, however, still open.

The basic question is why such a striking correlation between actional classes and morphemes exists and why the acquisition process follows this path. One possible answer may be provided by appealing to the notion of *prototype*. As is well known, this notion assumes that linguistic categories have an internal structure, with some members being more central (or prototypical) than others (Taylor 1989, among others). Thus, a prototype may be conceived as a cluster of prototypical semantic properties which are preferably correlated with certain morphological markings. Following this approach, Shirai and Andersen (1995:758) have claimed that "children acquire a linguistic category starting with the prototype of the category, and later expand its application to less prototypical cases". This developmental process, although with some variation, has also been observed in L2 learners. However, this claim is only a hypothesis about the learner's mental grammar. The theoretical implications of the notion of prototype for acquisition should be made explicit by asking where prototypes come from. Are they in children/learners' minds from the beginning? Needless to say, the question is particularly intriguing and crucial to L1 (see Bertinetto and Noccetti 2006).

But first let me discuss some research results achieved under the prototype hypothesis. In a systematic study of a learner corpus of Italian (available at the Department of Linguistics of the University of Pavia, see Andorno 2001),[7] I have claimed that the prototype approach may account for the acquisition and use of

[7] In the Pavia corpus learners are located at different levels of proficiency (Giacalone Ramat 2003c). Needless to say, the data discussed in this paper belong to low level proficiency learners, "basic variety" or "post-basic variety".

tense/aspect morphology in learner Italian (Giacalone Ramat 1995, 2003c). When the features of telicity, perfective aspect and past tense are congruent, the acquisition is facilitated (according to the "principle of selective association"). Indeed, learners of Italian start using past tense morphemes (namely auxiliary + past participles or bare participles for some learners) with telic verbs, and the same learners use the imperfective past forms first with stative verbs like *essere* "be", or activities:

(9) dopo due settimane noi ritornato Asmara
after two weeks we come back-PAST.PTCP to Asmara
"after two weeks we came back to Asmara"
Banca dati, MK, L1 Tigrinya, 03

(10) io ero piccolo millenovecentosetantatre Itiopia
I be.IMPF.1SG a little child in 1973 in Etiopia
"I was a little child in 1973 in Etiopia"
Banca dati, MK, L1 Tigrinya, 05

Admittedly, the developmental sequence may not be so linear for all learners, or even for the same learner, because of various intervening factors that it is not possible to specify here.

The prototype account, as formulated by Andersen (1991, etc.), was first applied to the emergence and use of past tense inflections. The association between telic predicate/perfective aspect/past tense appears to be encoded into the grammatical category "past tense" earlier than the parallel association between atelic predicate/imperfective aspect/present tense.[8]

To be sure, a "Progressive prototype" has also been proposed (Andersen and Shirai 1996:557–58, Shirai 2002) and its possible internal structure has been sketched, with "ongoing activity" as its core meaning. This version of the progressive prototype may be useful in languages in which learners can find overt progressive marking and thus establish a form-meaning association, as in English or Japanese (Shirai 2002), but is unable to account for the acquisition of languages lacking overt obligatory marking, like German, or Italian or French, which use periphrases or lexical expressions, or allow simple present or imperfect to convey progressive meaning. As an alternative, I suggest treating the Present as a prototype category. This means that in learner Italian the present would be the counterpart of the past prototype. The meaning components of the "Present Prototype" would presumably be atelicity and stativity associated with imperfective aspect (as shown above in (8)). The major uses of the present in native Italian do indeed fit with this claim (Bertinetto 1986).[9] The present emerges early in second learner varieties (and

[8] Quoting Dahl's (1985:78) characterization of the prototype for perfective, Andersen (2002:80) adds: "Thus, although Past is a Tense category and Perfective is an Aspect category, the two appear to have almost the same prototype".

[9] As shown by Bertinetto (1986:342ff), the present is a polyfunctional category. Although its prototypical aspectual value is imperfective, it may exhibit a number of perfective uses, like the present "pro futuro" (*vengo stasera* "I'm coming in the evening") or the "attitudinal" meaning (*Michele parla russo* "Michael speaks Russian") or the "atemporal" meaning (*da qui si vedono le montagne* "One can see the mountains from here").

also in children). On closer inspection, however, some data on the acquisition of L2 Italian turn out to be problematic with respect to this claim.

Learner productions do not provide any clear-cut evidence for the emergence of a "present prototype", because of the polyfunctionality of the present, which appears to be an unmarked category in the initial stages of acquisition. In learner Italian, a number of cases are found in which the present has a past temporal reference:

(11) /I/ lei non parlava italiano quando è arrivato?
"didn't you speak Italian when you arrived?"
\WZ\ sì sì appena in Idalia + non + non capisco, idalia tutto
yes yes as soon as in Italy not understand-PRES.1SG Italian all
++ poi ++ vai + vado in libllia + complale un liblo +
then go-PRES.2SG go-PRES-1SG in bookshop buy-INF a book
chiama + lingua per ai/ + la lingua idaliana per i stranieri +++
call-PRES.3SG language for the language Italian for the foreigners
in casa + ogni giolno + stasela + in casa + studiare ++ (xxx) due anni
at home every day this evening at home study-INF two years

"when I arrived in Italy, I didn't understand Italian, then I went to a bookshop and bought a book, it was called "Italian for foreigners" and I've been studying it at home every day in the evening for two years"

(Banca dati WZ, L1 Chinese, 01)

This is a complex text in terms of its temporal structure: past time reference is provided by knowledge of the situation (the learner is talking about his first experiences when he arrived in Italy), but is not marked morphologically. The learner first produces a stative predicate *capisco*, formally in the present, then a telic predicate always in the present referring to a perfective event in the past, *vado in libllia + complale un liblo*, then uses the infinitive for an activity with habitual aspect (*ogni giolno + stasela + in casa + studiare ++ (xxx) due anni*: the intended habitual meaning is conveyed through lexical means: *ogni giolno + stasela*).

The second example from the same learner provides corroborating evidence:

(12) Tu vivevi in campagna o in città?
"did you live in the country or in town"?
\WZ\ vivo in un grande città esciancra
live-PRES.1SG in a big city (Shangai)

"I used to live in a big city, Shangai"
WZ 03

This event is temporally interpretable thanks to the context: the learner is referring to his life in China. Both examples show a correlation between a durative predicate (*vivo*) and a stative predicate (*capisco*) with imperfective aspect, which is a congruent association, but the temporal reference would require imperfective past marking. In this example, as is usual in the early stages of the acquisition of Italian (Giacalone Ramat 1995), the present is a temporally unmarked form. On the other

hand, *vado in libllia* is a telic predicate, to which, according to the predictions of the prototype account, the past perfective morpheme should be readily assigned by learners. Still, we find a present form which, in native Italian, would be incompatible with the temporal reference (*appena in Idalia*).

Thus, in learner languages the emergence of prototypes is delayed because learners are (not yet) able to make all appropriate distinctions, in particular they do not master the formal marking of temporal reference. First they use an unmarked form which looks like the present (in some cases the infinitive is also used as unmarked form, see *studiare* above). Learners take more time to tease out the components of the "present prototype" than those of the past perfective.

After about twenty years since the prototype notion began to be applied to acquisition, not all problems have been solved. While the general outline of this explanation is not in dispute, empirical data have revealed a greater complexity than previously expected and, consequently, some implications of the prototype account have been called into question.

The role of the input has been discussed for both L1 and L2. It has been pointed out that in many languages native speakers too exhibit a particular distribution of tense and aspect markers with certain classes of verbs, thus the preferred association of telicity, perfective aspect and past tense is already present in the language the child or the L2 learner are exposed to. Consequently, the predictions made by the prototype model for the acquisitional path are, strictly speaking, weakened, or even made redundant because the input alone would account for the learner behavior. One might argue that children or L2 learners simply imitate the input.

However, the observed convergence with the input does not seem to be a sufficient explanation of learner behavior. Learners do not simply imitate, because, if this were the case, there should be no discrepancy with the input. Still, a significant deviation from native speakers can be observed in learners who are more restricted and rigid in their (early) use of morphological forms marking tense/aspect distinctions. In L1 acquisition too, children appear to move gradually, in their use of tense/aspect morphology, towards their caretakers' speech, although more research is required on child directed speech to assess this question (Bertinetto and Noccetti 2006). In such development, usage frequency is presumably an important factor: learners hear less frequently occurrences of stative verbs in the past perfective form (or telic verbs in the past imperfective form). But, again, the frequency argument alone does not explain the patterning of development.

The input effect had already been noted by Shirai and Andersen (1995), Andersen (2002), and labeled "The Distributional Bias Hypothesis". Andersen's suggestion is that the model offered by native speakers only reinforces a combination of features learners independently discover. According to Andersen (2002:81), learners are "cognitively predisposed" to notice in the input, and to learn to use, grammatical inflections (or auxiliaries) and verbs to associate them with. Andersen and Shirai seem to imply that the learner has an innate knowledge of the category of actionality or that the category with all its semantic distinctions is somehow "mature" in his/her mind. Bertinetto and Noccetti (2006) and Bertinetto et al. (in press) plead for a reformulation which does not take for granted the learner's knowledge of

the actionality category. Recent research (Giacalone Ramat and Rastelli, 2008) has taken issue with the assumption of the innate character of the semantic category of actionality. Learner data suggest that learners have to reconstruct the actional meaning and aspectual properties of predicates using a number of elements which may accompany the predicate. Learners tend first to apply a "global perspective" according to which verbal lexemes are underspecified with respect to temporal, aspectual and actional properties.

As an illustration consider the following example (from Giacalone Ramat and Rastelli 2008):

(13) \MK\ adesso io dico che cosa io fatto la scorsa settimana la scorsa settimana fino () fino a mercoledì ho imparato a al classe
"Now I'll say what I did last week last week until Wednesday I have learned in class"

Banca dati, MK, L1 Tigrinya, 09

In Italian *imparare* is a gradual completion verb (Bertinetto and Squartini 1995), i.e. a telic verb which indicates events made up of a sequence of successive, partial achievements. The past perfective *ho imparato* is not compatible with the culminative adverbial *fino a mercoledì*. The learner MK appears to ignore this restriction. Only in the last recording does MK correctly contrast *sto imparando* "I am learning" with *non ho imparato bene* "I did not learn well". This would suggest that telicity as well as other actional and aspectual properties of lexical verbs have to be learned: no innate predisposition can be assumed.

On closer inspection, the way the notion of prototype has been applied until now to acquisition does not account satisfactorily for learner behavior. From a purely descriptive point of view, it is fair to say that learners acquire first a prototype for the past perfective. However, there is apparently an early stage in the acquisition process in which no prototype has yet formed because the semantic-cognitive dimensions of temporality, aspectuality and actionality remain underspecified since learners do not yet have full mastery of the target morphology and/or do not know how to use it. It has to be borne in mind that the use of tense and aspect morphology is acquired in context, i.e. on the basis of how items are used in communication.[10]

To conclude, let us briefly go back to the issue of explanation. Why should the actionality category rather than, say, tense be prominent in the emergence of verbal morphology? A partial answer could be that tense in an utterance does not need any morphological codification, as is shown by those learner varieties in which temporal distinctions are conveyed by adverbials. Of course, in the case of L2 learners, a viable assumption could simply be that verb classes are known from

[10] On this subject, a criticism that can be leveled at Andersen and Shirai's valuable work is that they strongly rely on Vendler's classification. This classification has, historically, been fundamental and is a good starting point, but is essentially a lexical classification. As I said above, to describe learner data, theoretical proposals are needed which allow us to take larger units into consideration, such as the verb phrase, objects, and adverbials. That is, actionality is a compositional property (Verkuyl 1993).

the first language: according to Slobin (1985) an adult implicitly distinguishes a state from a telic predicate, even if his/her language lacks overt marking. As a matter of fact, however, typological research (Tatevosov 2002) has shown considerable cross-linguistic variability in the domain of actionality. The hypothesis of the native language filter for L2 acquisition does not fully account for learners' behavior.

In L1 acquisition, things may differ considerably since children have no previous grammatical knowledge.[11] One might assume, following Andersen, a kind of cognitive dominance, or relative ease of individuation, of certain event types in tense and aspect learning. However, this assumption is not needed. What we call actionality is a way of categorizing states of affairs which happen to exist in our world and of deciding what aspects of a situation are relevant, for instance the result achieved (*svuotare*, "empty", *rompere* "break") or the action performed (*passeggiare* "walk"). Category formation in acquisition probably emerges gradually, semantic distinctions are learned from recurrent situations and extracted from the input data learners hear around them, as is the case for other abstract cognitive operations. The mental operations involved in recognizing types of predicates should not be different from those used in other cognitive activities (such as the categorizing ability and the pattern finding ability). The crucial role played by such skills as pattern finding and analogy making in the development of child language has been pointed out by developmental psychologists (Tomasello 2003). This perspective also allows for the hypothesis that semantic dimensions (temporal reference, aspect, actionality) are initially underspecified. The L2 learner is in a similar situation with respect to the task of disentangling the various functions of verb inflections. In the perspective of acquisition the notion of prototype, which was elaborated independently of acquisition studies, is still relevant because it makes a clear prediction of initial underextension and gradual accretion (Taylor 1989:241). This is indeed what acquisitional data suggest, as we tried to show. The acquisition of the grammatical categories of tense and aspect involves a gradual assembling of the appropriate features. Prototypes are not available to the learner, but have to be inferred from the input. By contrast, Andersen and Shirai assume that prototypes can be directly found in morphological forms from the very beginning: "It is, therefore, possible to regard tense-aspect morphology as a prototype category consisting of good members and marginal members" (1996:556). To be sure, the existence of a prototype effect in L2 development is also confirmed by the data we analyzed (see above §4).

[11] Then the central question becomes: how does actionality develop in children? The answers have been different: a developmental psycholinguistic model, illustrated by Slobin (1985), assumes that the distinction between basic situation types, such as states and processes, is part of the cognitive maturation of all individuals. They would be conceptual notions of a universal nature which are distinguished on the basis of perceptual and cognitive properties. In the usage-based model, linguistic representations (categories) emerge gradually and in a piecemeal fashion depending on the type and token frequency with which children hear particular constructions (Tomasello 2003:142). Bertinetto and Noccetti (2006) start from the hypothesis of an inherently syncretic concept in which the temporal, aspectual and actional dimensions "appear inextricably intertwined".

The theoretical and methological issues discussed in this paper are an attempt at gaining a better understanding of the genesis of temporal and aspectual categories in learner languages. The findings of experimental studies appear more problematic than was initially assumed by Andersen and Shirai. They show that it is necessary to go beyond the "Primacy Aspect Hypothesis" as was formulated, and to assume that actionality, i.e. the semantic properties of predicates, is not available to learners, but is acquired, by taking into account the morphological and syntactic environments and discourse circumstances involved in the process of acquisition.

In future research the hypotheses and claims advanced so far should be tested more systematically using languages with different morphological patterns and different types of tense/aspect dominance.

This paper has wandered away from Greenberg. However, I think that these reflections follow the spirit of Greenberg's suggestion that L2 data are a valid source for linguistic theories, and also follow his inductive method for finding and validating universals. I am confident that this paper is one of the several ways in which L2 data can provide insights into theoretical issues.

References

Andersen, Roger W. (ed.). 1983. *Pidginization and creolization as language acquisition*, Rowley, MA: Newbury House.

Andersen, Roger W. 1991. Developmental sequences: The emergence of aspect marking in second language acquisition. In *Crosscurrents in second language acquisition and linguistic theories*, eds. Thom Huebner and Charles A. Ferguson, 305–24. Amsterdam/Philadelphia: John Benjamins.

Andersen, Roger W. 2002. The dimension of "Pastness". In *Description and explanation in the L2-acquisition of tense-aspect morphology*, eds. Rafael Salaberry and Yasuhiro Shirai, 79–105. Amsterdam: Benjamins.

Andersen, Roger W. and Yasuhiro Shirai. 1996. The primacy of aspect in first and second language acquisition: The pidgin/creole connection. In *Handbook of second language acquisition*, eds. William C. Ritchie and Tej K. Bhatia, 527–570. New York: Academic Press.

Andorno, Cecilia. 2001. *Banca dati di Italiano L2. Il progetto di Pavia.* Pavia: Dipartimento di Linguistica.

Antinucci, Francesco and Ruth Miller. 1976. How children talk about what happened. *Journal of Child Language* 3: 167–189.

Arends, Jacques, Pieter Muysken and Norval Smith (eds). 1994. *Pidgins and creoles. An introduction.* Amsterdam: Benjamins.

Aronoff, Mark. 1994. *Morphology by itself. Stems and inflectional classes*. Cambridge and London: The MIT Press.

Bernini, Giuliano. 1990. Lo sviluppo di paradigmi verbali nelle varietà elementari di apprendimento dell'italiano lingua seconda. In: *La temporalità nell'acquisizione di lingue seconde*, eds. Anna Giacalone Ramat and Giuliano Bernini, 81–101. Milano: FrancoAngeli.

Bernini, Giuliano. 2003. Learner varieties and language types: The case of indefinite pronouns in non-native Italian. In: *Typology and second language acquisition*, eds. Anna Giacalone Ramat, 83–123. Berlin: Mouton de Gruyter.

Bertinetto, Pier Marco. 1986. *Tempo, aspetto e azione verbale nel verbo italiano. Il sistema dell'indicativo*. Firenze: Accademia della Crusca.

Bertinetto, Pier Marco, Alessandro Lenci, Sabrina Noccetti, and Maddalena Agonigi in press. Metodi quantitativi nello studio dell'acquisizione delle strutture tempo-aspettuali. In *Linguistica e modelli tecnologici di ricerca*, eds. Ferrari Giacomo et alii, Proceedings of the XL International Conference of the Società di Linguistica Italiana (Vercelli, September 21–23, 2006).

Bertinetto, Pier Marco, and Sabrina Noccetti 2006. Prolegomena to ATAM acquisition. Theoretical premises and corpus labeling. *Quaderni del Laboratotio di Linguistica* 6/2006, Pisa: Scuola Normale Superiore.

Bertinetto, Pier Marco, and Mario Squartini. 1995. An attempt at defining the class of Gradual Completion Verbs. In *Temporal reference, aspect and actionality*. Vol. I: *Semantic and syntactic perspectives*, eds. Bertinetto, Pier Marco et al., 11–26. Torino: Rosenberg & Sellier.

Bickerton, Derek. 1977. Pidginization and creolization: language acquisition and language universals. In *Pidgin and Creole Linguistics*, eds. Albert Valdman, 49–69. Bloomington: Indiana University Press.

Bickerton, Derek. 1981. *Roots of language*. Ann Arbor: Karoma.

Chini, Marina. 2002. Ordres marqués et perspective du locuteur en italien L2. *Revue française de linguistique appliquée* VII: 107–127.

Chini, Marina. 2003. Morfologia del nome. In: (2003c) ed. Anna Giacalone Ramat, 37–69.

Comrie, Bernard. 1989. *Language universals and linguistic typology*, 2nd ed. Oxford: Blackwell.

Croft, William. 2001. *Radical construction grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, William. 2003. *Typology and universals*, 2nd ed. Cambridge: Cambridge University Press.

Cuzzolion, Pierluigi, Ignazio Putzu, and Paolo Ramat. 2006. The Indo-European Adverb in diachronic and typological perspective. *Indogermanische Forschungen* 111: 1–38.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Ferguson, Charles A. 1991. Currents in second language acquisition and linguistic theories. In eds. Thom Huebner and Charles A. Ferguson, 425–434.

Giacalone Ramat, Anna. 1992. Grammaticalization processes in the area of temporal and modal relations. *Studies in Second Language Acquisition* XIV: 297–322.

Giacalone Ramat, Anna. 1994. Il ruolo della tipologia linguistica nell'acquisizione di lingue seconde. In *Italiano lingua seconda/lingua straniera. Atti del XXVI Congresso della Società di Linguistica Italiana*, eds. Anna Giacalone Ramat and Massimo Vedovelli, 27–43. Roma: Bulzoni.

Giacalone Ramat, Anna. 1995. Tense and aspect in learner Italian. In *Temporal reference, aspect and actionality*, vol. II, *Typological perspectives*, eds. Pier Marco Bertinetto, Valentina Bianchi, Östen Dahl, and Mario Squartini, 289–307. Torino: Rosenberg & Sellier.

Giacalone Ramat, Anna. 1999. Le strategie di collegamento tra proposizioni nell'italiano di germanofoni. Una prospettiva di tipologia funzionale. In *Grammatik und Diskurs/Grammatica e discorso. Studi sull'acquisizione dell'italiano e del tedesco/Studien zum Erwerb des Deutschen und des Italienischen*, a cura di Norbert Dittmar and Anna Giacalone Ramat, 13–54. Tübingen: Stauffenberg.

Giacalone Ramat, Anna. 2002. How do learners acquire the classical three categories of temporality? In eds. Rafael Salaberry and Yasuhiro Shirai, 221–248.

Giacalone Ramat, Anna. 2003a. (ed.). *Typology and second language acquisition*. Berlin: Mouton de Gruyter.

Giacalone Ramat, Anna. 2003b. Gerunds as optional categories in second language learning. In (2003a), ed. Anna Giacalone Ramat, 181–220.

Giacalone Ramat, Anna (ed.). 2003c. *Verso l'italiano. Percorsi e strategie di acquisizione*. Roma: Carocci.

Giacalone Ramat, 2003d. Morfologia dell'italiano L2: difficoltà e strategie di sinofoni. In *Italiano L2 di cinesi. Percorsi acquisizionali*, a cura di. Emanuele Banfi, 11–24. Milano: FrancoAngeli.

Giacalone Ramat, Anna and Stefano Rastelli 2008. Learning actionality: an investigation of L2 Italian data. In *Empirische Forschung und Theoriebildung. Beiträge aus der Soziolinguistik,Gesprochene-Sprache-Forschung und Zweitspracherwerbsforschung. EineFestschrift für Norbert Dittmar zum 65. Geburtstag*, eds. Bernt Ahrenholz, Ursula Bredel, Wolfgang Klein, Martina Rost-Roth and Romuald Skiba, 239–250. Bern: Peter Lang.

Givón, Talmy. 1990. *Syntax. A Functional-typological Introduction*, vol. I. Amsterdam: Benjamins.
Givón, Talmy. 1995, *Functionalism and grammar*. Amsterdam: Benjamins.
Greenberg, Joseph H. 1960. A quantitative approach to morphological typology of language. *International Journal of American Linguistics* 26: 178–194 [first published in: Spencer R.F. (ed.), 1954. *Perspectives in Anthropology. Papers in Honor of W.D. Wallis*. Minneapolis: University of Minnesota Press, 192–220. Italian translation in Ramat, Paolo (ed.). 1976. *La tipologia linguistica*, 115–154. Bologna: Il Mulino].
Greenberg, Joseph H. 1966. Some universals of grammar with particular reference to the order of meaningful elements. In *Universals of language*, ed. Joseph H. Greenberg, 73–113. Cambridge, MA: The MIT Press (trad. it. in P. Ramat (a cura di), *La tipologia linguistica*, 115–154. Bologna: Il Mulino 1976).
Greenberg, Joseph H. 1991. Typology/universals and second language acquisition. In eds. Thom Huebner and Charles A. Ferguson, 37–43.
Hammarberg, Bjoern and Koptjevskaja-Tamm, Maria. 2003. Adnominal possession: combining typological and second language perspectives. In ed. Anna Giacalone Ramat 2003a, 125–179.
Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford: Oxford University Press.
Heidelberger Forschungsprojekt 'Pidgin Deutsch'. 1976. Untersuchungen zur Erlernung des Deuschen durch ausländische Arbeiter. Germanistisches Seminar der Universität Heidelberg.
Holm, John. 2000. *A Introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
Huebner, Thom and Charles A. Ferguson (eds). 1991. *Crosscurrents in second language acquisition and linguistic theories*. Amsterdam/Philadelphia: John Benjamins.
Hyltenstam, Kenneth. 1986. Markedness, language universals, language typology and language acquisition. In *First and second language acquisition processes*, ed. Carol Pfaff, 55–78. Cambridge, MA: Newbury House.
Keenan, Edward. L. and Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, VIII, 1, 63–99 (trad. it. Accessibilità del sintagma nominale e grammatica universale. In *Introduzione alla tipologia linguistica*, eds. Sonia Cristofaro and Paolo Ramat, 107–139). Roma: Carocci 1999.
Keenan, Edward L. and Comrie, Bernard. 1979. Data on the noun phrase accessibility hierarchy. *Language* 55: 333–351.
Klein, Wolfgang. 1986. *Second language acquisition*. Cambridge: Cambridge University Press.
Klein, Wolfgang. 1991. SLA theory: prolegomena to a theory of language acquisition and implications for theoretical linguistics. In eds. Thom Huebner and Charles A. Ferguson, 169–194.
Klein, Wolfgang and Perdue Clive. 1992. *Utterance structure. Developing grammars again*. Amsterdam: Benjamins
Klein, Wolfgang and Clive Perdue. 1997. he Basic Variety (or: Couldn't Natural Languages be Much Simpler?) *Second Language Research* 13: 301–347.
Kouwenberg, Silvia and Peter L. Patrick (eds). 2003. Reconsidering the role of SLA in Pidginization and Creolization. Special issue of *Studies in Second Language Acquisition*, vol. 25, n.2.
McWhorter, John H. 1998. Identifying the Creole Prototype. Vindicating a typological class. *Language* 74: 788–818.
McWhorter, John H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5,2/3: 125–166.
Muysken, Pieter and Norval Smith (eds). 1986. *Substrata versus universals in creole genesis*. Amsterdam: Benjamins.
Muysken, Pieter and Tonjes Veenstra. 1994. Universalistic approaches. In *Pidgins and creoles. An introduction*, eds. Jacques Arends, Pieter Muysken, and Norval Smith, 121–134. Amsterdam: Benjamins.
Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, Mass.: The MIT Press.
Ramat, Paolo. 1999. Linguistic categories and linguists' categorizations. *Linguistics* 37: 157–180 (revised edition in *Pagine Linguistiche*, 61–88. Bari: Laterza, 2005).
Romaine, Suzanne. 1988. *Pidgin and creole languages*. London: Longman.
Rutherford, William E. (ed.). 1984. *Language universals and second language acquisition*. Amsterdam: Benjamins.

Salaberry, Rafael and Shirai, Yasuhiro (eds.). 2002. *Description and explanation in the L2-acquisition of tense-aspect morphology*. Amsterdam: Benjamins.
Schumann, John. 1978. *The pidginization process*. Rowley, Mass.: Newbury House.
Shirai, Yasuhiro. 2002. The prototype hypothesis of tense-aspect acquisition in second language. In eds. Rafael Salaberry and Yasuhiro Shirai, 455–478.
Shirai, Yasuhiro and Roger W. Andersen 1995. The acquisition of tense-aspect morphology: a prototype account. *Language* 71: 743–762.
Shirai, Yasuhiro and Yumiko Nishi. 2003. Lexicalization of aspectual structures in English and Japanese. In ed. Anna Giacalone Ramat, 2003a, 266–290.
Slobin, Dan I. (ed.). 1985. *The crosslinguistic study of language acquisition*, vol. I, *The data*; vol. II, *Theoretical issues*. Hillsdale (NJ): Erlbaum.
Slobin, Dan I. 1996. From 'thought and language' to 'thinking for speaking'. In *Rethinking linguistic relativity*, eds. John J. Gumperz and Steven C. Levinson, 70–96. Cambridge: Cambridge University Press.
Tatevosov, Sergej 2002. The parameter of Actionality. *Linguistic Typology* 6: 317–401.
Taylor, John. 1989. *Linguistic categorization. Prototypes in linguistic theory*. Oxford: Clarendon Press.
Tomasello, Michael. 2003. *Constructing a language. A usage-based theory of language acquisition.* Cambridge, Mass.: Harvard University Press.
Vendler, Zeno. 1967. Verbs and Times. *Linguistics in Philosophy*. Ithaca, NY: Cornell University Press: 97–121.
Verkuyl, Henk J. 1993. *A theory of aspectuality. The interaction between temporal and atemporal structures.* Cambridge: Cambridge University Press.
Voeikova, Maria D. and Dressler Wolfgang U. (eds). 2002. *Pre- and protomorphology.* Muenchen: Lincom Europa.
WALS. 2005. *The world atlas of language structures*, eds. Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie. Oxford: Oxford University Press.

Language Index

Subject Index


sharing is caring!!!
BMN.06