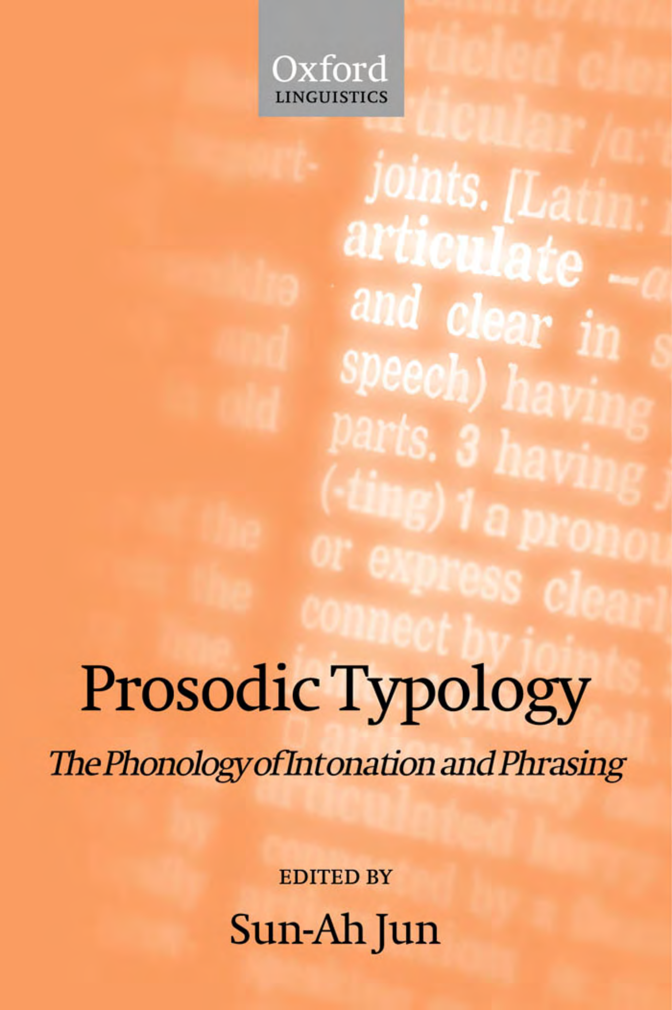
Oxford
LINGUISTICS



Prosodic Typology

The Phonology of Intonation and Phrasing

EDITED BY

Sun-Ah Jun

Prosodic Typology

The Phonology of Intonation and Phrasing

This page intentionally left blank

Prosodic Typology

The Phonology of Intonation and Phrasing

Edited by

SUN-AH JUN

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published in the United States
by Oxford University Press Inc., New York

First published 2005
First published in paperback 2006

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India
Printed in Great Britain
on acid-free paper by
Biddies Ltd., King's Lynn

ISBN 0-19-924963-6 978–0–19–924963–3
ISBN 0-19-920874-6 (Pbk.) 978–0–19–920874–3 (Pbk.)

109 8 7 6 54321

Contents

Preface

This project of editing a book on prosodic typology started in 1998. Researchers working on intonation in the framework of Autosegmental-Metrical phonology were invited to present their work at the Intonation: Models and Transcription Workshop, a satellite meeting of the 14th International Congress of Phonetic Sciences in San Francisco, California, in 1999. All papers presented at the workshop, except for one on French, are included in the current volume. I am grateful to all the participating authors for their patience and their timely contributions to this long-term project.

I am also grateful to many people who have been involved in the process of editing this book and organizing the Intonation Workshop: to Mary Beckman for encouraging me with the idea of editing a book and organizing the workshop, and for serving as a discussant at the workshop; to Julia Hirschberg and Stefanie Shattuck-Hufnagel for their help in serving as discussants at the workshop and for their moral support in the editing of this book; to those who helped at the registration desk at the Intonation Workshop (Taehong Cho, Cecile Fougeron, Sahyang Kim, Misun Seo, Minjung Son, and Motoko Ueyama); to Amy Schafer for her help in writing the book proposal; to the reviewers of individual chapters and the book (Robert S. Bauer, Judith Bishop, Gösta Bruce, Caroline Féry, Janet Fletcher, Matthew Gordon, Martin Grice, Bob Ladd, Kikuo Maekawa, Tomas Riad, Chilin Shih, Jennifer Venditti, and four anonymous reviewers); to Mary Baltazani for her help in editing an earlier version of this book; to Mary Baltazani and Henry Tehrani for preparing the earlier version of the CD-ROM; to Illam Park for shaping the final format of the CD-ROM; to Barry Griner for proofreading the manuscripts and checking the formatting of references and all the sound files in the CD-ROM; and to Pat Keating for her encouragement and support especially when I was not very mobile for several months in 2000. Finally, I would also like to thank the linguistics editor of Oxford University Press, John Davey; the production editors in linguistics, Henry Miller, Jacqueline Smith, and Stuart Fowkes; the copy-editor, Christine Ranft; and their linguistics assistant, Karen Morgan, for their guidance, patience, and encouragement. This work was partially supported by a UCLA Senate grant.

Sun-Ah fun

List of Contributors

Amalia Arvaniti (University of California, San Diego, USA)

Cinzia Avesani (Institute for Cognitive Sciences and Technologies of the National Research Council, Italy)

Mary Baltazani (University of Ioannina, Greece)

Stefan Baumann (U. des Saarlandes, Germany)

Mary E. Beckman (Ohio State University, USA)

Ralf Benzmiiller (G Data Software AG, Germany)

Judith Bishop (University of Melbourne, Australia)

Gösta Bruce (Lund University, Sweden)

Marjorie K. M. Chan (Ohio State University, USA)

Mariapaola D'Imperio (CNRS, France)

Janet Fletcher (University of Melbourne, Australia)

Svetlana Godjevac (HS Technologies, Inc., San Francisco, USA)

Matthew K. Gordon (University of California, Santa Barbara, USA)

Esther Grabe (University of Oxford, UK)

Martine Grice (University of Cologne, Germany)

Carlos Gussenhoven (Radboud University, Nijmegen, and Queen Mary, University of London)

Julia Hirschberg (Columbia University, USA)

Tsan Huang (State University of New York, Buffalo, USA)

Sun-Ah Jun (University of California, Los Angeles, USA)

Ok Joo Lee (Ohio State University, USA)

Shu-hui Peng (National University of Kaohsiung, Taiwan)

Michelina Savino (University of Bari, Italy)

Stefanie Shattuck-Hufnagel (MIT, USA)

Chiu-yu Tseng (Academia Sinica, Taiwan)

Jennifer J. Venditti (Columbia University, USA)

Paul Warren (Victoria University of Wellington, New Zealand)

Wai Yi P. Wong (Ohio State University, USA)

1

Introduction

Sun-Ah fun

This book is an edited volume of multiple chapters, each of which contributes to establishing prosodic typology. It includes descriptions of the intonation and the prosodic structure of thirteen typologically different languages based on the same theoretical framework, the Autosegmental-Metrical (AM) model of intonational phonology (Bruce 1977; Pierrehumbert 1980; Ladd 1983, 1996; Gussenhoven 1984; Liberman and Pierrehumbert 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Pierrehumbert and Hirschberg 1990). The languages included vary geographically, ranging from European languages (English, German, Greek, Dutch, Serbo-Croatian, Italian, and Swedish), Asian languages (Japanese, Korean, Cantonese, and Mandarin), a Native American Indian language (Chickasaw), and an Australian aboriginallanguage (Bininj Gun-wok, also known as Mayali). They also vary in the type and the degree of lexical specifications of prosody. Some have lexical stress (e.g. English, Greek), lexical pitch accent (e.g. Japanese), or tone (e.g. Mandarin, Cantonese), while others have both stress and pitch accent (e.g. Swedish, Chickasaw) or none of these (e.g. Korean). Among these, eleven languages are also described in the prosodic transcription system known as ToB! (Tones and Break Indices; Beckman and Hirschberg 1994; Beckman and Ayers-Elam 1997): English, German, Greek, Italian, Serbo-Croatian, Japanese, Korean, Mandarin, Cantonese, Chickasaw, and Bininj Gun-wok.

This book thus demonstrates how a single approach to a prosodic model and transcription system can be applied to typologically different languages, even when the languages vary considerably in their prosodic features as well as their morphosyntactic structures. Describing the prosody of multiple languages in the same theoretical framework would make it significantly easier for us to observe universal and language-specific prosodic features as well as greatly advance our understanding of prosodic typology.

The AM model of intonational phonology adopted in this book assumes that intonation has a phonological organization (Pierrehumbert 1980; see Ladd 1996 for the criteria). It describes intonation as a sequence of distinctive tonal units (High and Low and their combinations) forming a hierarchical prosodic structure. The intonation contour is represented linearly by an autosegmental string of tones, which are aligned with a specific syllable or with a specific location in a phrase, marking the prominence relations among the words and the prosodic groupings of an utterance, i.e., metrical and prosodic structure. The phonological representation of tones is mapped onto a phonetic representation through phonetic realization rules, and as in segmental phonology, both the phonological representation and the phonetic realization of prosodic features are language specific.

In addition to the tonal patterns defined by intonational phonology, a prosodic structure of a language is also defined by the degree of juncture between two adjacent words. This is reflected in the ToBI transcription system. ToBI transcribes, at a minimum, distinctive tonal events ('To' forTone) and the perceived degree of disjuncture between any two words ('BI' for Break Index). The tone labels represent the intonational phonology of each language, and the categories of the break index, a numerical value, represent the metrical hierarchy of the prosodic groupings of each language. Prosodic groupings are often marked by intonation or duration, or both. Therefore, the tonal event and the break index information together describe the prosodic system of a language (see Chapter 2 for a detailed description ofToBI).

Because the prosodic model assumed in ToBI is a phonological model, not a phonetic one, the ToBI system of one language is not appropriate for describing the prosodic system of other languages or even other dialects ofthe same language. Thus, the ToBI transcription system is different from the INTSINT (INternational Iranscription System for INTonation) transcription system proposed by Hirst and Di Cristo (1998), which transcribes pitch movements as a sequence of static points or pitch targets, using symbols (e.g. '→' for a pitch point at the same level as the immediately preceding one; '↑' for a pitch point relatively higher than the preceding one). As stated in Hirst and Di Cristo (1998), INTSINT is equivalent to a narrow phonetic transcription of segments and is used for gathering data on languages whose inventory of tonal patterns has not been established and for transcribing multiple languages based on the surface fo pattern of the utterances. The ToBI transcription system further differs from INTSINT in that the INTSINT transcribes only pitch information while ToBI transcribes pitch and other prosodic information including the degree of juncture (see Chapter 2, Section 2.6, for a comparison between ToBI and other phonetic transcription systems).

As described in detail in Chapter 2, ToBI, in its origin, was the transcription conventions used to cover Mainstream American English (Silverman *et al.* 1992; Pitrelli *et al.* 1994). However the principles and annotation conventions of American English ToBI have been applied to other English varieties (e.g. Glasgow ToBI by Mayo *et al.* 1997) and a number of other languages as described in this book, thus making it possible to compare prosodic systems across dialectsllanguages using the same terminology. The original ToBI has been renamed MAE_ToBI (Mainstream American English ToBI) in Chapter 2 following the general practice of putting the initials of the language in front of'ToBI' (e.g. GToBI for German ToBI, K-ToBI for Korean ToBI, and LToBI for Japanese ToBI).[1]

The original ToBI conventions specify four layers oflabelling, aligned with speech signals: words, tones, break indices, and miscellaneous information. But the information that can be labelled is not fixed. As shown in several ToBI models in this book, the number and types of tiers (where prosodic categories and information relevant to prosody are labelled) reflect the prosodic system specific to each individual language and the interests of the research group.[2]

The benefit of transcribing prosodic information is not confined to finding out the prosodic patterns of the language and refining the model. It also helps researchers to discover the relationship between prosody and sub-areas of the grammar of a language (e.g. syntax-prosody mapping, prosody in semantics and sentence processing, prosodic features reflecting a discourse structure and sociolinguistic variations). Thus, in addition to establishing prosodic typology, developing a ToBI system for a language would contribute to our understanding of the role of prosody in the grammar. Recently, researchers have started to investigate the role of intonation (defined phonologically) in pragmatics and discourse structure (e.g. Ward and Hirschberg 1985; Hirschberg and Ward 1995; Swerts and Hirschberg 1998; Venditti 2000; Park 2003), in semantics (e.g. Steedman 1991,2000; Jun and Oh 1996; Roberts 1996; Büring 1997, 1999; Baltazani 2002), in sentence processing and word segmentation (e.g. Speer *et al.* 1993; Schafer *et al.* 1996; Fodor 1998, 2001, 2002; Hirose 1999; Kjelgaard and Speer 1999; Schafer and Jun 2002; Jun 2003; Kim 2004), in prosodic phonology (e.g. Jun 1993, 1998; Post 2000; D'Imperio

[1] The usage of ToBI has been generalized in some linguistics literature as another name for a phonological model of intonation based on Pierrehumbert's model, or a name for an American model of intonation in contrast to European or British models, i.e., ToBI model = a Pierrehumbert style intonation model. This generalization has probably been caused by the fact that the 'To' part (= intonation) of ToBI has been used more often than the 'BI' part (= indexing prosodic boundary strength) in the discussion of intonation.

[2] The original ToB! used the software *xwaves* (Entropies) to transcribe ToBI labels. However there is no mandatory software for transcribing ToBI labels. In this book, six ToBIs used *xwaves,* four used *Pitchworks* (Scicon R&D, *http://www.sciconrd.com).* and one used *EMU* (see Ch. 9 for details).

and Fivela 2004), in phonetics (e.g. Beckman and Edwards 1990, 1994; Edwards *et al.* 1991; de Jong *et al.* 1993; Fougeron and Keating 1997; Fougeron and Jun 1998; Fougeron 1999; Cho and Keating 2001; Keating *et al.* 2004), in language acquisition (e.g. Jannedy 1997; Jun and Oh 2000; Choi and Mazuka 2002), and in sociolinguistics (e.g. Kang 1996; Park 2003).

Though linguists, psychologists, and speech scientists have become increasingly aware of intonational phonology and the ToB! transcription for English and other languages, a complete description of the ToB!s of these languages had not been published. It is believed that this book will serve as a good reference and guide to researchers as they start or build both a phonological model of intonation and a ToB! system for other languages from scratch, and will further encourage them to analyse more languages in the same framework. The book is designed for advanced undergraduate and graduate students and researchers in linguistics, psychology, language teaching, speech science, and engineering. It is expected to be read by linguists who are interested not only in the phonetics and phonology of intonation but also in the role of prosody in the sub-areas of grammar. It will also be read by speech scientists interested in speech synthesis and recognition since the transcription system is closely related to what has been developed and used by people in the speech industry. Finally, this book is also for language teachers-by comparing the intonational categories and their realizations in the target languages, they can pin down the sources of prosodic interference and transfer.

The organization of the book is as follows. Chapter 2, written by three of the system's first developers, introduces the history and background of the original ToB! system; identifies the principles for designing both the original and the general ToB! prosodic transcription systems; and discusses the extensions of the basic ToB! tiers. The thirteen subsequent chapters present the intonational phonology of thirteen typologically different languages. The first ten of these chapters describe the prosodic system of an individual language, while the last three of these chapters (Italian, English, and Swedish) describe the prosodic differences of dialects within a language. With the exception of Dutch, the chapters describing the prosody of an individual language first introduce the intonational phonology of that language and then its ToB! system, increasing coherence among chapters and facilitating cross-linguistic comparisons. The chapter on Dutch describes only the intonation system, based on Gussenhoven's model, another AM model of intonational phonology.

The ten chapters on the prosody of an individual language are ordered from well-studied European languages (Chapters 3-6; German, Greek, Dutch, Serbo-Croatian) to Asian languages (Chapters 7-10; Japanese, Korean, Mandarin, Cantonese) to fieldwork languages (Chapters 11-12; Chickasaw, Bininj Gun-wok). Within the European languages, stress languages (German, Greek, Dutch) are followed by a pitch accent language, Serbo-Croatian, which is then followed by an Asian pitch accent language, Japanese. Among the Asian languages, the Mandarin chapter comes before the Cantonese chapter because the description of the latter is compared with the former. Bininj Gun-wok is presented in the last chapter of the ten individual languages because this chapter discusses prosodic differences among its dialects, similar to the topic of the subsequent three chapters (Chapters 13-15), which discuss how to transcribe intonation patterns of different dialects of the same language in the AM framework. Of the three languages on dialect typology, Swedish intonation is described based on Bruce's AM model (Bruce 1977). Together, this book illustrates the issue of typology on a different scale, a dialect typology versus a language typology in terms of prosody. To help the readers to appreciate the description of intonation in each chapter, sound files (in a 'waves' format) corresponding to pitch tracks in each chapter are stored in a CD-ROM, which is attached to the back cover of this book.

The last chapter of this book compares the eleven ToB! systems described in the book, and discusses the flexibility and the extensions of the ToB! model. This chapter also proposes a model of prosodic typology based on twenty-one languages (the thirteen languages described in this book and eight other languages described in the AM model). The prosodic similarities and differences across languages are captured by two categories, i.e., Prominence, and Rhythmic/Prosodic Unit, with each category being further divided into lexical and postlexicallevels. It is hoped that the intonational phonology and transcription systems of more languages will be included in the second edition of this book to enrich the data for prosodic typology. We still have a long way to go before establishing a complete picture of a prosodic typology, but we hope that this book at least gives us a good start in this direction.

REFERENCES

BALTAZANI, M. (2002), 'Quantifier Scope and the Role ofIntonation in Greek', Ph.D. dissertation (UCLA).

BECKMAN, M. E. and AYERS-ELAM, G. (1997), *Guidelines for ToBI Labeling.* Ms. (Ohio State University) [http://ling.ohio-state.edu/~tobi].

—— and EDWARDS, J. (1990), 'Lengthenings and Shortenings and the Nature of Prosodic Constituency', in J. Kingston and M. Beckman (eds.), *Between the*

Grammar and Physics of Speech: Papers in Laboratory Phonology I (Cambridge: Cambridge University Press), 152-78.

BECKMAN, M. E. and HIRSCHBERG, J. (1994), 'The ToB! Annotation Conventions?', ms. Ohio State University.

— — and PIERREHUMBERT, J. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook,* 3: 255-309.

BRUCE, G. (1977), *Swedish Word Accents in Sentence Perspective* (Lund: Gleerup).

BURING, D. (1999), 'Topic', in P. Bosch and R. van der Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives* (Cambridge, UK: Cambridge University Press), 142-65.

CHO, T. and KEATING, P. (2001), 'Articulatory and Acoustic Studies of Domain-initial Strengthening in Korean', *Journal of Phonetics,* 29: 155-90.

CHOI, Y. and MAZUKA, R. (2002), 'Young Children's Use of Prosodic Cues in Sentence Processing', a talk presented at the 15th Annual meeting of CUNY Conference on Human Sentence Processing (CUNY Graduate School and University Center).

DE JONG, K., BECKMAN, M. E., and EDWARDS, J. (1993), 'The Interplay between Prosodic Structure and Coarticulation', *Language and Speech,* 36: 197-212.

D'IMPERIO, M. and FIVELA, G. (2004), 'How many Levels of Phrasing? Evidence from two Varieties of Italian', in J. Local, R. Ogden, and R. Temple (eds.) *Papers in Laboratory Phonology VI* (Cambridge: Cambridge University Press).

EDWARDS, J., BECKMAN, M. E., and FLETCHER, J. (1991), 'Articulatory Kinematics of Final Lengthening', *Journal of the Acoustical Society of America,* 89: 369-82.

FOUGERON, C. (1999), 'Articulatory Properties of Initial Segments in Several Prosodic Constituents in French', *UCLA Working Papers in Phonetics,* 97: 74-99.

— — and JUN, S.-A. (1998), 'Rate Effects on French Intonation: Phonetic Realization and Prosodic Organization', *Journal of Phonetics,* 26/1: 45-70.

— — and KEATING, P. (1997), 'Articulatory Strengthening at Edges of Prosodic Domains', *JASA,* 101/6: 3728-40.

FROTA, S. (1998), 'Prosody and Focus in European Portuguese', Ph.D. Dissertation (University of Lisbon).

GUSSENHOVEN, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

HIROSE, Y. (1999), 'Resolving Reanalysis Ambiguity in Japanese Relative Clauses', Ph.D. dissertation (The City University of New York, New York).

HIRSCHBERG, J. and WARD, G. (1995), 'The Interpretation of the High-rise Question Contour in English', *Journal of Pragmatics,* 24: 407-12.

HIRST, D. and DI CRISTO, A. (1998), 'A Survey of Intonation Systems', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems: Survey of Twenty Languages* (Cambridge: Cambridge University Press).

JANNEDY, S. (1997), 'Acquisition of Narrow Focus Prosody', Proceedings of the GALA '97 Conference: Language Acquisition, Knowledge Representation & Processing.

TUN, S.-A. (1993), 'The Phonetics and Phonology of Korean Prosody', Ph.D. dissertation (Ohio State University, Columbus, Ohio). [Published in 1996 by Garland Publishing Inc., New York: NY]

— — (1998), 'The Accentual Phrase in the Korean Prosodic hierarchy', *Phonology,* 15/1: 189-226.

— — (2003), 'Prosodic Phrasing and Attachment Preferences', *Journal of Psycholinguistic Research,* 32/2: 219-49.

— — and OH, M. (1996), 'A Prosodic Analysis of Three Types of Wh-Phrases in Korean', *Language and Speech,* 39/1: 37-6l.

— — , — — (2000), 'Acquisition of Second Language Intonation', *ICSLP* (Beijing, China), 4: 76-9.

KANG, H.-S. (1996), 'Acoustic and Intonational Correlates of the Informational Status of Referring Expressions in Seoul Korean', *Language and Speech, 39(4):* 307-40.

KEATING, P., CHO, **T.,** FOUGERON, C. and Hsu, c.-S. (2004), 'Domain-initial Strengthening in Four Languages', in *Laboratory Phonology* 6 (Cambridge: Cambridge University Press). [Also appeared in *UCLA Working Papers in Phonetics* 97: 139–51.]

KIM, S. (2004), 'The Role of Prosodic Phrase in Word Segmentation', Ph.D. dissertation (UCLA).

KJELGAARD, M. M. and SPEER, S.-R. (1999), 'Prosodic Facilitation and Interference in the Resolution of Temporary Syntactic Closure Ambiguity', *Journal of Memory and Language,* 40: 153-94.

LADD, R. (1983), 'Phonological Features ofIntonational Peaks', *Language,* 59: 721-59.

— — (1990), 'The Metrical Representation of Pitch Register', in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I* (Cambridge, UK: Cambridge University Press), 35-57.

— — (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

LIBERMAN, M. and PIERREHUMBERT, J. (1984), 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aronoff and R. Oerhle (eds.), *Language Sound Structure* (Cambridge, MA: MIT Press), 157-233.

MAYO, c., AYLETT, M. and LADD, D. R. (1997), 'Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI', Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications (Athens, Greece), 18-20 September 1997, 231-4·

NESPOR, M. and VOGEL, 1. (1986), *Prosodic Phonology* (Dordrecht: Foris).

OSTENDORF, M., PRICE, P. and SHATTUCK-HuFNAGEL, S. (1995), 'The Boston University Radio News Corpus'. Boston University Technical Report No. ECS-95-00l.

PARK, M.-J. (2003), 'The Meaning of Korean Prosodic Boundary Tones', Ph.D. dissertation (UCLA).

PIERREHUMBERT, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (MIT) [distributed by lULC 1987].

PIERREHUMBERT, J. and BECKMAN, M. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

--and HIRSCHBERG, J. (1990), 'The Meaning of Intonational Contours in the Interpretation of Discourse', in P. Cohen, J. Morgan, and M. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271-311.

PITRELLI, J. F., BECKMAN, M. and HIRSCHBERG, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToB! Framework', Proceedings of International Conference on Spoken Language Processing (pp. 123-6), Yokohama.

POST, B. (2000), 'Tonal and phrasal structures in French intonation' (Doctoral Dissertation, (University of Nijmegen), The Hague: Holland Academic Graphics.

PRICE, P., SHATTUCK-HUFNAGEL, S. and FONG, C. (1991), 'The Use of Prosody in Syntactic Disambiguation', *Journal of the Acoustical Society of America, 90/6:* 2956-70.

ROBERTS, C. (1996), 'Information Structure in Discourse: Towards an Integrated Formal Theory of Pragmatics', in J. H. Yoon and A. Kathol (eds.) *OSUWPL Volume* 49, *Papers in Semantics,* Ohio State University Department of Linguistics, 91-136. Available on the Semantics Archive: http://www.semanticsarchive.net/

SCHAFER, A. J., CARTER, J., CLIFTON, C. and FRAZIER, 1. (1996), 'Focus in Relative Clause Construal', *Language and Cognitive Processes,* 11 (1/2): 135-63.

— — and JUN, S.-A. (2002), 'Effects of Accentual Phrasing on Adjective Interpretation in Korean', in M. Nakayama (ed.), *East Asian Language Processing,* Stanford: CSLI, 223-55·

SELKIRK, E. (1986), 'On Derived Domains in Sentence Phonology', *Phonology Yearbook,* 3: 371-405·

SILVERMAN, K., BECKMAN, M. E., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C., PRICE, C., PIERREHUMBERT, J. and HIRSCHBERG, J. (1992), 'ToB!: A Standard for Labeling English Prosody', *Proceedings of International Conference on Spoken Language Processing* (pp. 867-70). Banff.

SPEER, S. R., CROWDER, R. G. and THOMAS, 1. (1993), 'Prosodic Structure and Sentence Recognition', *Journal ofMemory and Language,* 32: 336-58.

STEEDMAN, M. (1991), 'Structure and Intonation', *Language,* 67/2: 260-96.

— — (2000), 'Information Structure and the Syntax-phonology Interface', *Linguistic Inquiry,* 31: 649-89.

SWERTS, M. and HIRSCHBERG, J. (1998), 'Prosody and Conversation: An Introduction', *Language and Speech,* 41(3-4): 229-33.

VENDITTI, J. (2000), 'Discourse Structure and Attentional Salience Effects on Japanese Intonation', Ph.D. dissertation. (Ohio State University).

WARD, G. and HIRSCHBERG, J. (1985) 'Implicating Uncertainty', *Language,* 61: 747-76.

WIGHTMAN, C., SHATTUCK-HUFNAGEL, S., OSTENDORF, M., and PRICE, P. (1992), 'Segmental Durations in the Vicinity of Prosodic Phrase Boundaries', *JASA, 91/3:* 1707–17.

2

The Original ToBI System and the Evolution of the ToBI Framework

Mary E. Beckman, Julia Hirschberg, and Stefanie Shattuck-Hufnagel

2.1. INTRODUCTION

The term ToBI has come to be used in two different ways. Originally, it was the name of an annotation system, developed in the period 1991 to 1994, which was designed for use in labelling intonation and prosody in databases of spoken Mainstream American English (Beckman and Hirschberg 1994). Very quickly, however, it also came to refer to a general framework for the development of prosodic annotation systems in other varieties of English (e.g. Mayo *et ai.* 1997 [Glasgow]) and in other languages (e.g. Grice *et ai. 1996* [German]; Venditti 1997 [Japanese]). In this chapter, we will try to identify the essential properties of a ToBI framework annotation system by describing the development and design of the original ToBI conventions. In this description, we will overview the general phonological theory and the specific theory of Mainstream American English intonation and prosody that we decided to incorporate in the original ToBI tags. We will also state the practical principles that led us to make the decisions that we did.

Before we begin, however, we should explain a practical terminological decision. Although the original ToBI for Mainstream American English (MAE) is the most completely developed of the ToBI-framework systems, and also the system most completely tested by use, the development of ToBI-framework systems for other languages makes this dual usage increasingly awkward. Therefore, we will adopt the following convention for distinguishing the two uses in this chapter. We will reserve the unmodified term 'ToBI' for the developmental framework, and use the prefixed term 'MAE_ToBI' for the original system.

The chapter is organized as follows. Section 2.2 briefly chronicles how the MAE_ToBI system came into being. The purpose of this chronicle is to bring out the practical principles that guided the system's development. Section 2.3 briefly describes the consensus account of English intonation and prosody on which the MAE_ToB! system is based. In such a necessarily abbreviated account, we will not be able to begin to do justice to the more than eighty years of observation and instrumental research that have made the intonation systems of MAE and its close relative, standard Southern British English (including RP), among the best understood in the world. Section 2.4 catalogues the different components of a MAE_ToB! transcription and lists the salient rules which constrain the relationships between different components. This section also expands upon the theoretical foundations and practical consequences of adopting the general structure of multiple labelling tiers, and particularly the separation of the labels for tones from the labels for indexing prosodic boundary strength. Section 2.5 then describes some of the extensions of the basic ToB! tiers that have been adopted by some sites. This section also compares our decisions about the number of tiers and about inter-tier constraints with the analogous decisions for some of the other ToB! systems described in this book. Section 2.6 discusses the status of the symbolic labels relative to the continuous phonetic records that are also an obligatory component of the MAE_ToB! transcription. In particular, we describe the status of the Tones tier relative to the fundamental frequency record, and contrast this status to the epistemological claims implicit or explicit in some other transcription frameworks. Section 2.7 then closes by listing several open research questions that we would like to see addressed by MAE_ToB! users and the larger ToB! community.

2.2. THE BEGINNINGS OF MAE_ToB!

MAE_ToB! was developed in a series of four meetings, with delegates from a large number of sites, representing several disciplines. Participants in the workshop included engineers who wanted to train automatic speech recognition systems and build better text-to-speech systems, psychologists who wanted to investigate the relationship between prosody and human language processing, computer scientists who wanted to build better dialogue models and speech generation systems, phoneticians who wanted to test theories about tone association and alignment, and so on. This diversity of aims was also reflected in the diversity of sites for the meetings. The initial Prosodic Transcription Workshop (1-2 August 1991) was organized by Victor Zue at the MIT Laboratory for Computer Science. The Second Prosodic Transcription Workshop (5-6 April 1992) was organized by Kim Silverman at

NYNEX Science and Technology, Inc. The Third Prosodic Transcription Workshop (17-21 June 1993) was organized by Mary Beckman at the Department of Linguistics, Ohio State University. And the Fourth Prosodic Transcription Workshop (5-6 August 1994) was organized by Mari Ostendorf at the Department of Electrical, Computing, and Systems Engineering, Boston University.

Each of these four meetings was termed a 'workshop' and the term was appropriate. At each of them, we arrived prepared to work hard, and we did work hard to agree on what phenomena we wanted the conventions to cover and to decide on how to cover these phenomena. Also, before each workshop, participants prepared a set of exercises, transcribing a common set of utterances, to focus discussion at the meeting. The transcriptions prepared for the first meeting gave us an immediate basis for comparing existing transcription conventions across sites, a comparison which showed us that we had considerable grounds for consensus already in the group. That is, delegates from four of the eleven sites represented at the first workshop happened to have transcribed tone contours using the same autosegmental model of intonation (Pierrehumbert 1980; Beckman and Pierrehumbert 1986), and informal comparison of these transcriptions at the workshop suggested a high degree of agreement among them. Also, delegates from another three sites had transcribed prosodic grouping using the same system of numerical indices to boundary strength (Price *et al.* 1991), and had compared their transcriptions prior to the meeting in order to be able to present actual numbers on inter-transcriber reliability. The transcription exercises prepared for the second meeting similarly served as the basis for our first group-wide inter-transcriber consistency check (Silverman *et al. 1992).* Moreover, thanks to Patti Price, we benefited from a meticulous record of the first meeting, and continued to make detailed minutes at subsequent meetings. These records allowed us to circulate a summary of the decisions made at each meeting immediately afterwards. They also served as the basis for successive drafts of the ToB! conventions, which were revised after each of the first three meetings before their final codification in Beckman and Hirschberg (1994). Between the second and the fourth meetings, we also developed a set of on-line training materials to accompany the draft annotation conventions and for use by new transcribers (Beckman and Ayers 1994). These training materials allowed us to recruit naïve transcribers, who had not attended any of the meetings, for the second inter-transcriber consistency check (Pitrelli *et al.* 1994).

To understand some of our decisions at these workshops, it is useful also to know what prompted us to convene them in the first place. The immediate impetus was the example of the Penn TreeBank project (Marcus *et al.* 1993). By agreeing on a common core of syntactic labels, the Natural Language Processing community had been able to develop a large online corpus of syntactically annotated English text. This corpus has been used for exploring aspects of syntactic theory (e.g. Srinivas and Joshi 1999), for testing models of sentence processing (e.g. MacDonald *et al.* 1994), and for improving the performance of syntactic parsers (e.g. Collins 1999; Charniak 2000). We were inspired by the TreeBank example to try to find an analogous set of consensus tags for intonation and prosody. In the short term, we wanted a similar tool that would allow researchers at different sites to share in the work of developing a large pool of prosodically transcribed online speech databases for a broad range of uses in speech science and technology. In the longer term, we wanted to provide a common vocabulary so that researchers at different sites could interpret each other's data and contribute complementary analyses and extensions of a common core of methods and datasets. There are practical principles to extrapolate from four aspects of this original purpose and of the work that we did to achieve that purpose.

First, the MAE_ToB! system did not spring out of thin air. Rather, it is based on long history of studies of English intonation, stress, and phrasing. Moreover, a sizeable group of researchers in the original MAE_ToB! community were versed enough in various aspects of this history to have provided some basis for consensus even at the first meeting. (Section 2.3 reviews the antecedents of this consensus.) The lesson for the larger speech community is that any new set of ToB! conventions for another language needs to reflect a fairly broad and well-grounded understanding of the intonational and prosodic grammar of the language. Ideally, the conventions will be based on a large and long-established body of research in intonational phonology, dialectology, pragmatics, and discourse analysis for the targeted language variety, but at the least, they should be based on rigorous analyses of the intonational phonology. Where established analyses are available for only a subset of the phenomena that users want to label, the development of a ToB! framework system can help formulate the relevant questions for further research, but system development should not run too far ahead of knowledge.

Second, the initial MAE_ToB! group was large enough and diverse enough-approximately 25-30 people attended each workshop-to allow us to pool expertise of various types. The system that eventually emerged was designed to cover only a subset of the prosodic features which we wanted to identify-namely, those that required hand labelling and which we collectively felt we knew enough about to label consistently. Some other features, such as the segmental makeup of each word or the location of its primary lexical stress, could be generated automatically using resources such as online dictionaries; thus, these were not explicitly included as part of the MAE_ToB! system. Still other features, such as phrasal stress patterns or subtle variation in the extent of pitch rises or falls that might be related to discourse-prominence relationships within and across intonational phrases, appeared to require more basic research before they could be labelled consistently; it was hoped that the creation of large labelled corpora could facilitate this study. The lesson we derive from our discussions of what to include in the MAE_ToB! conventions is that ToB! conventions should be efficient. One should not waste transcriber time by asking the transcriber to symbolically mark phenomena that can be extracted from the signal or derived from online resources automatically. (We review several examples, one involving location of main lexical stress, in Section 2.5.)

Third, the MAE_ToB! system was intended for use by an even larger community of end users, with a wide variety of interests and theoretical convictions. It is based on a broad consensus, which involved some amount of compromise on the part of every delegate to the workshops. The lesson to extrapolate for the larger ToB! framework is that a viable ToB! system needs a suitably large and diverse group of users who have agreed to act communally to develop and adopt the system as a community-wide standard. One corollary principle is that the conventions should be easy enough to teach that their use is not limited to a few experts to do the transcription and to themselves train apprentice labellers. Therefore, there must be a freely available manual for teaching the system to new transcribers, with many recorded examples of transcribed utterances graded from easy to difficult. Another corollary principle is that the conventions need to be used and maintained consistently across transcription sites. Therefore, in the course of developing a ToB! framework system, there must be rigorous tests of inter-transcriber consistency, and there should be agreed-upon centres for maintaining the standard with periodic rechecks and evaluation of any proposed revisions. A third corollary principle is that mechanisms must be provided for customizing transcriptions to particular needs without compromising the common core. (We elaborate on this point in Section 2.5.)

Fourth, the building of transcription conventions was a strikingly iterative process. It involved much discussion and often impassioned argument, interspersed with actual transcribing by all delegates of actual recorded utterances provided by all of the sites. In each iteration, we had a chance to see what kinds of prosodic phenomena were important to others, and how a proposed change in the transcription conventions would affect other transcribers' ability to capture what they heard in the signal and what they wanted to capture for their research. Moreover, the discussion of problems and proposed changes to the conventions was always grounded in the examination of actual speech signals. For example, to argue for making a distinction between two pitch accent categories that had been tagged the same way in the previous draft of the conventions, one at least had to articulate clearly what aspects of the signal supported the proposed distinction and show that the difference was based on more than one pair of examples. Ideally, one also could mimic the tunes on other texts and invoke a recognizable difference in felicitous contexts of use. The lesson for the larger community is that a good ToB! system is not simply a transcription system. It is also a tool for observing the signal and communicating one's observations to the larger community in a common language. A ToB! transcription never replaces a permanent record of the speech signal with a symbolic record; rather, it seeks to integrate a symbolic commentary with the data upon which it is based. A related lesson is that some phenomena of interest, such as phrasal pitch range variation, can be represented by continuous measures that are derived directly from the signal in conjunction with the symbolic labels. We elaborate on these points in Sections 2.4 and 2.6. First, however, we describe the consensus model of MAE intonation and prosody on which the original ToB! system was built.

2.3. THE ANTECEDENTS OF MAE_ToB!

The MAE_ToB! system is based on a consensus model that makes five salient claims about intonation and prosodic structure in the language. First, the prosodic pattern for an utterance can be projected onto separate tiers representing conceptually independent structural types. In particular, the intonation contour can be represented linearly by an autosegmental string of *tones*, whereas the metrical hierarchy of intonational phrases and lower-level prosodic groupings should be represented hierarchically, for example by a numerical *break index* value for the perceived degree of disjuncture between any two words. Second, the intonation contour is decomposed into relatively high and relatively low pitch levels: H versus L tones. These pitch levels are static targets in paradigmatic contrast with each other; 'relatively low' means low relative to the local phrasal *pitch range*, rather than low relative to the nearest pitch peak or plateau. This means that, for example, there can be simple L* and H* pitch accents in contrast with each other, as well as rising and falling accents, as in a dynamic tone model. Third, the local pitch range is determined by a variety of effects, such as phrasal prominence relationships or the occurrence of *downstep* (a compression of phrasal pitch range that reduces a 'downstepped' non-initial H target and all following H tones within the phrase) and *upstep* (a raising of the phrasal pitch range beginning at a H-phrase accent). These effects are specified independently of the tone level, so that a H tone in one part of the intonation contour for an utterance can be lower than a L tone elsewhere in the same utterance. Fourth, the tones for any phrase are distinguished functionally either as being *edge tones* or as being affiliated with *pitch accents*. The absolute pitch value of a tone depends on its function as well as on its position; for example, a L tone that defines the beginning of a L+H* rising pitch accent can be higher than a L% tone at the following intonational phrase boundary. The function of a tone also determines its timing relative to the autosegmental projection of consonants and vowels; a pitch accent is aligned to the segments of the relevant stressed syllable whereas an edge tone is aligned to the segments at the relevant phrase boundary. Fifth, there are contrastively H versus L edge tones at two levels of intonational phrasing, associated with two different degrees of juncture or boundary strength (i.e., the *intermediate phrase* versus the *intonational phrase).* Moreover, the lower-level edge tone-the *phrase accent-is* aligned to affect the pitch contour beginning immediately after the last tone target of the accent that is aligned to the syllable with *nuclear stress* (the most prominent accented item in the intermediate phrase). Thus, the phrase accent defines the beginning of the post-nuclear *tail.*[1] This decomposition of the contour beginning at the syllable with nuclear stress means that the intonation contour over the pre-nuclear *head* can be described in terms of the same inventory of pitch accent types available for the nuclear accent.

The immediate antecedents of the MAE_ToB! model are Pierrehumbert (1980) and Beckman and Pierrehumbert (1986) for the decomposition of the intonation contour into functionally distinct groups ofH versus L tone; Ladd (1983) for the treatment of downstep; and Price *et al.* (1991) and Wightman *etal.* (1992) for the treatment ofjuncture. However, all aspects of the model are also grounded in a long history of research on intonation and prosody in the language. For example, both the claimed relationship between the intonation contour and the sense of disjuncture or phrasing, and the notion of the pitch accent, predate Pierrehumbert's (1980) grammar of MAE tunes by several decades. The MAE intonation system is closely related to that of standard

[1] In accounts by British language teachers and phoneticians before the 1980S, the 'nucleus' of an intonation contour was modelled as a holistic dynamic tonal event governing the part of the contour beginning at the most stressed syllable. When this nucleus occurs far from the end of the contour, then the pitch pattern on material after the nuclear stress is called the 'tail'. The general shape of the intonation contour over accented syllables before the nucleus is then the 'head'.

Southern British English (SBE), which is one of the most studied in the world. Our modern understanding of the inventory of SBE intonation patterns traces its roots to astute observations by teachers of English as a foreign language beginning in the first decades of the last century (Palmer 1922; Armstrong and Ward 1926) and studies of large corpora in the 1960s (Halliday 1967; Crystal 1969) and later (e.g. Gussenhoven 1984). Ladd (1980) summarized important points of consensus among the then current competing models of SBE intonation, and pointed out the relationship between these models of SBE and Bolinger's (1958, 1964) observations of intonation patterns in MAE. These consensus properties include (1) the use of pitch changes (dynamic tones) rather than pitch levels (tone targets) to identify an abstract representation of the tune, independent of the associated text; (2) the connection drawn between tune and phrasing for at least one, and often for two levels of grouping (d. Trim 1959); (3) the notion of pitch accent as encoded, for example, in Kingdon's (1939) *tonetic stress marks;* and (4) the specification of a distinct inventory of tone patterns for the nuclear versus pre-nuclear accents (d. the review of treatments of the head in Ladd 1980). The second and third of these properties are the earlier antecedents for the claimed link between edge tones and phrasing, and for the notion of pitch accent.

Note, however, that the first and fourth of the properties that Ladd listed are not shared by the model of American English intonation that is encoded in the MAE_ToB! conventions. Instead, MAE_ToB! follows Pierrehumbert (1980) in adopting a tone target model rather than a dynamic tone model, and also a decomposition of the intonation contour starting at the nucleus that unifies the inventory of nuclear and pre-nuclear accent types.

The unified inventory for nuclear and pre-nuclear pitch accents builds on Bruce's (1977, 1982) seminal work on Swedish accent and intonation. In Bruce's model the pitch contour for an utterance is decomposed into a sequence of independent tone targets associated with different parts of the text. Every content word has an associated word accent and every intonational phrase has a phrase accent. Word accents are anchored at the stressed syllable in each successive word, the phrase accent is inserted into the tone string immediately after the lexical pitch accent of the word bearing the sentence stress (which is not necessarily the last content word), and boundary tones are anchored at the phrase boundaries. Variation in word tone shapes can then be neatly explained (and handily captured in speech synthesis) by modelling the effects of neighbouring tones (e.g. *undershoot* when two tones are crowded together) and of phrasal position (e.g. the word accent tones are successively reduced by downstep after the phrase accent) independent of the word tone specification. Pierrehumbert (1980) proposed a similar decomposition of American English intonation contours into tones contributed by the pitch accents, and two types of edge tones: boundary tones proper (which are anchored only at the intonational phrase edge) and phrase accents (which are anchored both to the end of the intermediate phrase and after the nuclear pitch accent in the phrase). With this decomposition of the pitch contour around and after the nuclear accented syllable, Pierrehumbert was able to unify the treatment of accent in the head and in the nucleus, and to account for the shape of the tail. The model also resolves a fuzziness inherent in dynamic tone accounts of alignment when a contour tone is 'stretched out' over different numbers of syllables under different conditions of stress and phrasing. The fall-rise nucleus, for example, does not simply flatten out to yield a shallower fall and shallower rise when the post-nuclear tail is longer than a few syllables. Rather, the low inflection point stretches out into a plateau, so that the fall remains steep and anchored at the stressed syllable and the rise remains steep and anchored at the phrase end. A level tone model specifies this alignment pattern by decomposing the dynamic shape into a sequence of simple tones that are aligned to specific prosodic landmarks. The fall and the rise of a fall-rise nucleus, for example, can be described as a sequence of H* pitch accent, L- phrase accent, and H% boundary tone, which are anchored invariantly to the appropriate prosodic domain however many syllables there are after the sentence stress. (See Pierrehumbert 1980, 2000; Liberman and Pierrehumbert 1984; and Pierrehumbert and Beckman 1988, for a discussion of this and other empirical advantages of the level-tone model as compared to a dynamic-tone model.)

The tone target analysis adopted in the MAE_ToB! model also resolves problems with earlier models of MAE intonation that posited more than two tone levels. For example, Bolinger (1951) criticized Trager and Smith's (1951) four-level analysis for its failure to differentiate continuous pitch range variation from categorical intonational contrast. (See Pierrehumbert 1980 or Pierrehumbert 2000 for a summary of this and other problems.) Pierrehumbert's solution was to posit tone targets, but only at two rather than four or five levels, and then to develop a more elaborated model of the relationship between tone target and backdrop pitch range in order to account for the observations of systematic phonetic differences among more than just the one relatively low ('L') and the one relatively high ('H') tone level. The elaborations of the model of the relationship between tone target and pitch range are of two types: positionally-conditioned local variation in the realization of the same tone (e.g. a 'L%' boundary tone target anchored at a phrase boundary typically is lower than the 'L' target in a 'L+H*' pitch accent) and more global variation in the backdrop pitch range (e.g. downstep reduces the pitch of a 'H' and all following 'H' targets until the end of the phrase).

Pierrehumbert's two-tone model of English intonation owes much to insights about tone patterns in African tone languages (e.g. Leben 1973; Anderson 1978), and has been adopted widely in the subsequent literature on intonation in English (e.g. Gussenhoven 1984) and in other languages (cl. Ladd 1996). In common with the Africanists, Pierrehumbert (1980) understood downstep to be triggered by a H L H alternation of tone levels within a phrase. In English, such an alternation only occurs when at least the first of the H tones is part of a pitch accent. In standard Japanese, downstep occurs specifically when a HL sequence is a pitch accent (see, *inter alia,* McCawley 1968; Pierrehumbert and Beckman 1988). An alternative understanding of downstep in English, therefore, is that it is triggered by the alternation of tone levels specific to a rising or falling pitch accent (see Beckman and Pierrehumbert 1986). A third alternative, proposed by Ladd (1990), is that downstep is a direct, iconic signal of reduced prominence for a later accent relative to an earlier accent. None of these accounts, however, has been tested against large corpora oflabelled speech. The MAE_ToB! conventions, therefore, adopt a more theory-neutral approach, following Ladd (1983). Specifically, downstep is marked explicitly on the first affected tone, using the downstep diacritic common in autosegmental treatments of African tone languages. Thus, 'L+!H*' signifies a rising pitch accent that is downstepped relative to an immediately preceding H tone target. While the immediate effect is that the peak value is intermediate between that of immediately preceding Hand L tones, this is not a mid (M) tone, but a H tone in a compressed pitch range. That is, the'!' in the tag also indicates the beginning of a stretch of speech in a compressed range, such that not only the downstepped !H target in the L+!H*, but all subsequent H tones will be lower relative to their 'expected' value had they occurred before the downstepped tone.

The treatment of juncture in Price *et al.* (1991) also has a long history. Essentially, it unifies earlier phonetic work on the relationship between juncture and syntax, and on discourse level effects on duration and pitch range (e.g. Lehiste 1960, 1975; O'Malley *et al.* 1973; Cooper 1976; Nakatani *et al.* 1981; Gee and Grosjean 1983), with the treatment of segmental sandhi effects in metrical phonology (e.g. Selkirk 1978; Nespor and Vogel 1986). The further insight achieved by comparing transcriptions at the first Prosodic Transcription Workshop in 1991 was that this treatment of juncture could also be linked to the treatment of intonational phrasing in Beckman and Pierrehumbert (1986) by identifying the perception of juncture at break index levels 3 and 4 as being cued in part by intonation contour shapes at the edges of Beckman and Pierrehumbert's (1986) intermediate phrase and intonational phrase, respectively. We also agreed that any perceived variation in boundary strength above the intonational phrase (BI = 4) could not be identified with the domains of prosodic effects such as the distribution of phrase tones, and therefore, that 'utterance' and 'paragraph' ends should not be marked (see Section 2.4). The transcription of tones following Beckman and Pierrehumbert (1986) and Ladd (1983) thus provides the *Tones* part of MAE_ToBI, and the transcription of perceived juncture at the ends of prosodic units in Price *et al.* (1991) provides the Break Indices part. In the next section we will review these and other obligatory parts of a MAE_ToBI record.

2.4. OVERVIEW OF THE MAE_ToB! CONVENTIONS

A full MAE_ToBI record of an utterance has at least six parts, listed in Table 2.1 and illustrated in Figures 2.1 through 2.3. Of these six parts, two are continuous phonetic records and four are symbol strings. The primary continuous phonetic record is an audio recording of the utterance. In the case of the utterances in the three figures, these are digital recordings on the CD that accompanies this book. The waveform in the top panel of each figure is a graphic representation of this recording. The other continuous phonetic record is some representation of the fundamental frequency (Fo) contour. This could be an analogue representation such as a narrow-band spectrogram, or a digital representation such as a string of numbers calculated by some Fo tracking algorithm. If the representation is of the latter form, it is

TABLE 2.1 The six obligatory parts of a MAE_ToB! record

Audio	An audio recording of the utterance in some form.
Fo	An electronic and/or paper record of the fundamental frequency contour.
Tones	An autosegmental transcription of the intonation contour; other tone related tags.
Words	An orthographic transcription of each word in the utterance, placed at the word's end, which is marked with a time index.
Break-Indices	Numeric index of the perceived degree of juncture after each orthographic word.
Misc	Markers for disfluencies, comments, and other miscellaneous events.

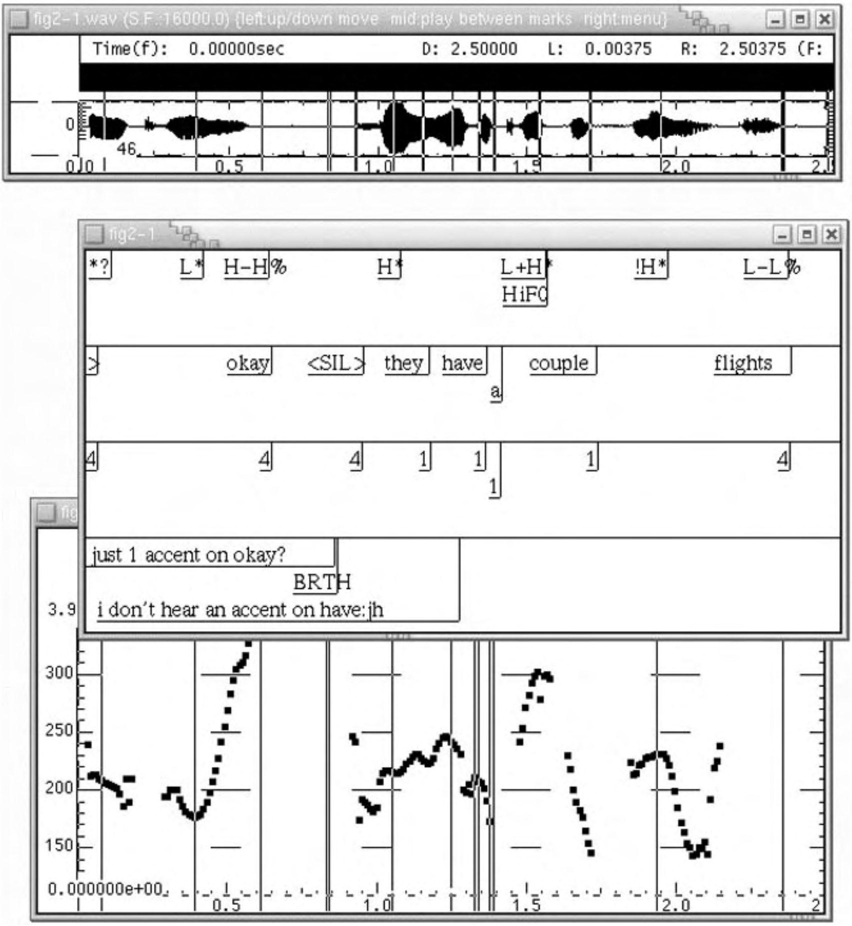


fig2-1.wav (S.F.:16000.0) {left:up/down move mid:play between marks right:menu}
Time(f): 0.00000sec D: 2.50000 L: 0.00375 R: 2.50375 (F:
0
46
0.0 0.5 1.0 1.5 2.0 2.
fig2-1
? L H-H% H* L+H* !H* L-L%
HiF0
> okay <SIL> they have a couple flights
4 4 4 1 1 1 1 4
just 1 accent on okay?
BRTH
i don't hear an accent on have:jh
3.9
300
250
200
150
0.000000e+00
0.5 1.0 1.5 2.0 2

FIGURE 2.1 Audio waveform, Fo contour, and MAE_ToBI xlabel windows for utterance *Okay... They have a couple flights.*

useful to provide also some graphical representation of the numbers, as in the bottom panel of each figure, which shows the output of an autocorrelation-based Fo tracking programme applied to the audio recording. The panel in between the waveform and the Fo contour in each figure shows the four obligatory symbol strings, ordered vertically in these displays starting with the tier of labels for *Tones* (i.e., a symbolic transcription of the intonation contour) at the top. A full list of the ten basic tonal morphemes and of the other Tones-tier labels is shown in Table 2.2. The two symbol strings just below the Tones tier are labels for all of the *Words* in the audio recording (i.e., an orthographic transcription) and labels of *Break Indices* (i.e., a number

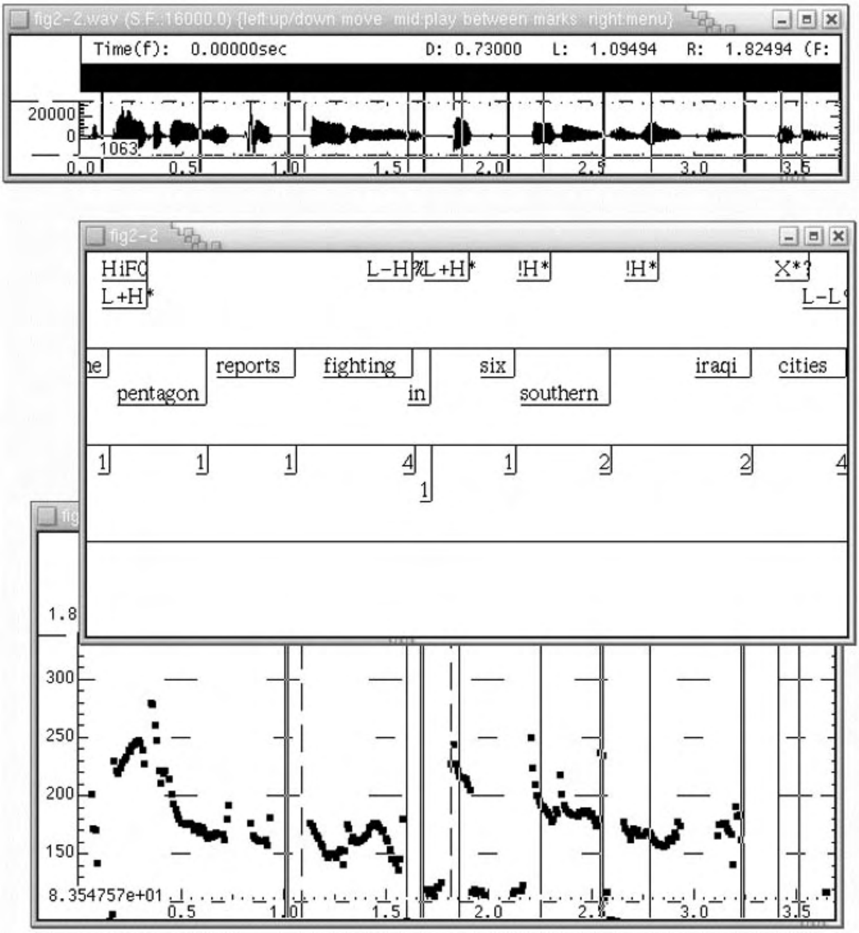


FIGURE 2.2 Audio waveform, Fo contour, and MAE_ToBI xlabel windows for utterance *The Pentagon reports fighting in six southern Iraqi cities.*

indicating the perceived degree of boundary strength) for each of the labels in the Words tier. The four[2] basic Break-Index values and the several diacritics and other labels for phenomena such as a marked prolongation that disrupts the intonation contour are given in Table 2.3. At the bottom of the panel

[2] Note that there are only four basic break index values, ordered from 0 to 4, with a 'hole' at 2. In the original Price *et al.* (1991) use of break indices, the value 2 represented a perceived boundary strength intermediate between a normal word boundary and a larger phrase boundary, and was used to mark a number of imprecisely-defined phenomena. The ToBI system restricts the use of this label to an explicit

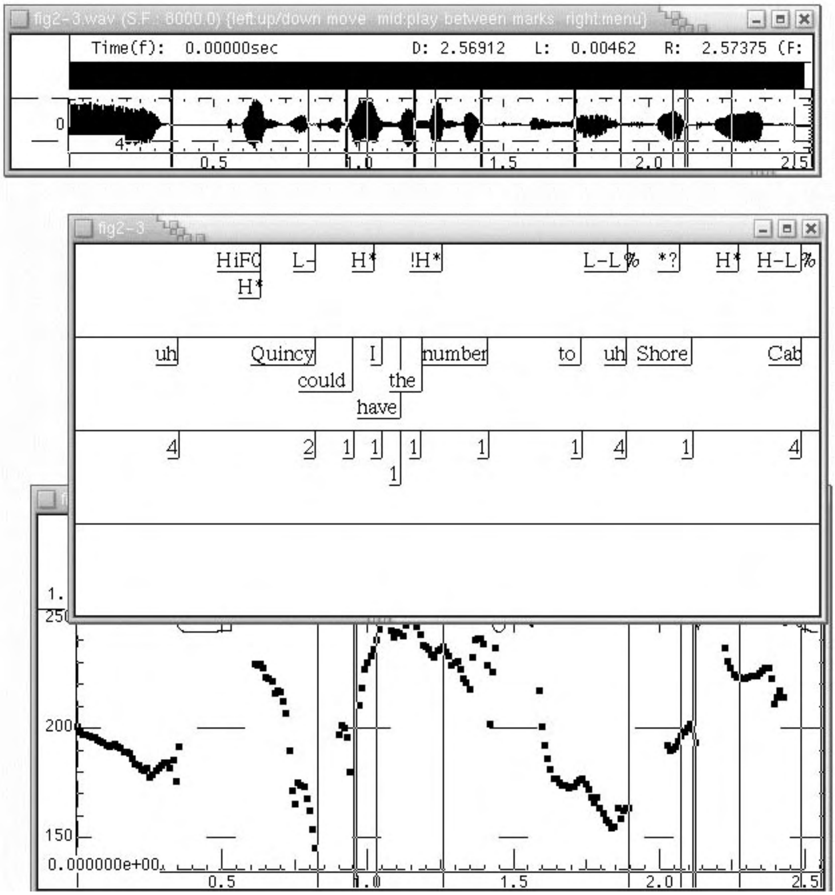


fig2-3.wav (S.F.: 8000.0) {left:up/down move mid:play between marks right:menu}
Time(f): 0.00000sec D: 2.56912 L: 0.00462 R: 2.57375 (F:
0
0.5
1.0
1.5
2.0
2.5
fig2-3
HiF0
H*
L-
H*
!H*
L-L%
*?
H*
H-L%
uh
Quincy
could
I
have
the
number
to
uh
Shore
Cab
4
2
1
1
1
1
1
1
4
1
4
1.
250
200
150
0.000000e+00

FIGURE 2.3 Audio waveform, F0 contour, and MAE_ToBI xlabel windows for utterance *Uhh… Quincy. Could I have the number to uh… Shore Cab?*

showing the symbol strings is the *Misc* tier of labels for events such as coughs or disfluencies—anything marking speech events of interest or questions or unusual configurations of labels on the other three symbolic tiers. (The Misc tier may optionally contain other labels, depending upon the aims of the

subset of these phenomena—namely, inter-word junctures where there is ambiguity between a 1 and a 3 either because there is a phrase tone without the duration lengthening appropriate to a 3, or a lengthening appropriate to a 3 but no phrase tone. This means that ToBI labels do not recognize a prosodic constituent comparable to Selkirk's (1995) 'Major Phrase' unless this is equated with Beckman and Pierrehumbert's (1986) tonally marked 'intermediate phrase'. Labellers who postulate and perceive a constituent boundary that is larger than a 'Prosodic Word' but smaller than the lowest intonationally marked constituent are encouraged to mark these events in a comments tier (see Section 2.5).

Table 2.2 The inventory of MAE_ToBI Tones-tier labels

Basic tones:	
phrase accents:	H- (!H-), L- (obligatorily placed at every BI = 3 and higher)
boundary tones:	H%, L% (obligatory at every 4) %H (marginal, at beginnings of some intonational phrases after pause)
pitch accents:	L*, H* (!H*), L+H* (L+!H*), L*+H (L*+!H), H+!H*
Other labels:	
downstep:	e.g. !H*, L+!H*, !H- (the ! diacritic marks the beginning of compressed pitch range)
uncertainty:	*?, -?, %? (uncertainty about occurrence); X*?, X-?, X%? (about tone type)
phonetic events transcribed in careful labelling:	< (delayed peak); HiF0 (maximum F0 associated with H of an accent within an intermediate phrase)
restart: %r (see the Misc tier)	

Table 2.3 The inventory of MAE_ToBI Break-Indices tier labels

Basic break index values:

0 (very close inter-word juncture)
1 (ordinary phrase-internal word end)
3 (intermediate phrase end, with phrase accent)
4 (intonational phrase end, with boundary tone)

Diacritics:

- (uncertainty)—e.g. 4- (intermediate between 3 and 4)
p (perceived hesitation)—1p for ‘cut-off’, 2p and 3p for ‘prolongation’

Tones-breaks mismatch:

2 (perceived 1 with unexpected tonal marker, or lengthening, etc., suitable for break index 3 or 4 without the phrase accent and/or boundary tone)

labelling project. See, for example, the Misc-tier labels in Figure 2.1 and the discussion in Section 2.6.)

Each of the labels on the symbolic tiers is an index to events observed in one or both of the continuous phonetic records, and this indexing function is accomplished by the time stamp associated with the label. For example, the time stamp for a label on the Words tier indexes the end of that word in the audio signal, and each label on the Break-Indices tier should share the same time stamp as the label for the immediately preceding word. Thus, in Figure 2.1, there are six word labels (discounting the' <SIL> '-see below), and each of these labels is aligned with a 1 or a 4 on the Break-Indices tier beneath it. Some of the labels on the Tones tier also inherit these time stamps. For example, the 'H-H%' sequence at the end of *okay* in Figure 2.1 is aligned with the time stamp marked by Break-Index 4. Other Tones-tier labels (e.g. the 'L*' preceding the 'H-H%' sequence in Figure 2.1) must be provided with their own independent time stamps that refer to the fundamental frequency record or to both the Fa and the audio recording. That is, the tags on the Tones tier are of three types, with slightly different alignment conventions. The three types oftone labels are (1) edge tone labels at each break index 3 and4; (2) pitch accent labels for each accented syllable within the intermediate phrase; and (3) two 'phonetic' labels at points where it is useful to extract times and fundamental frequency values in investigation of peak alignment and phrasal pitch range. These last two labels are the'<' tag that in careful labelling can be used to mark accent peaks which are not aligned with the relevant accented syllable, and the 'HiFo' label that helps gauge the pitch range for an intermediate phrase that contains at least one accent-related 'H' tone. For example, the utterance in Figure 2.1 contains two intermediate phrases, but only the second has a HiFo label, because the H tones in the first are both edge tones rather than accent tones. The placement of these two labels depends on the placement of accent labels, and will be described in more detail below (see also Section 2.6). By contrast, the edge tones depend on the Break-Indices tier (or the Words tier in some cases).

That is, the edge tones are the 'L-' and 'H-' phrase accents (marking the ends of all intermediate phrases), the 'L%' and 'H%' final boundary tones (marking the ends of all intonational phrases), and the '%H' initial boundary tone. The original MAE_TaB! decision that only the end of every word needs to be marked on the Words tier is related to the fact that every intermediate phrase must end with a phrase accent, and every intonational phrase must end with the phrase accent of the last intermediate phrase and an immediately following boundary tone, whereas the initial boundary tone is rather marginal in English. (Typically, the first well-defined tone target in an utterance is a pitch accent, on the first likely candidate syllable for accent, giving rise to the percept of stress shift if it is a syllable with lexically 'secondary stress'-cf., Shattuck-Hufnagel *et al.* 1994.) Thus, a phrase accent or a final boundary tone can simply inherit the time stamp for the break index that marks the end ofthe relevant intermediate phrase or intonational phrase. The '%H' initial boundary tone, on the other hand, must be aligned with the beginning of the phrase-initial word, and this will not already be marked in the Words tier, unless the transcriber has inserted a '<5IL>' or '#' label after each pause (see below).

The pitch accents include two accents in which the Fo remains low or falls to a lower level on the accented syllable ('L*' and 'H+!H*'), two accents in which the Fo remains high or rises to a peak on the accented syllable ('H*' and 'L+H*'), and a scooped accent ('L*+H') which has an Fo minimum within the accented syllable followed by an Fo peak. In careful labelling, these three sets of accent labels are placed differently with respect to the accented syllable. That is, minimally the label for a pitch accent should be placed somewhere within the syllable to which it is associated, so that the time stamp can identify the accented syllable when there is more than one candidate in a word. In careful transcriptions, however, further constraints are followed-namely, this time stamp is placed at the Fo minimum for 'L*', 'H+ !H*', and 'L*+ H', and at the Fo maximum for 'H*' and 'L+H*' so long as the maximum occurs within the accented syllable. If the maximum is later than the end of the syllable, the accent label is aligned to the amplitude peak within the accented syllable, and a '< ' label is placed at the actual Fo peak.

The '< ' can also be used to mark the Fo peak of 'L*+ H' (i.e. the Fo peak following L*) in such careful labelling, which is particularly useful if the aim is to investigate the phonetics of alignment. For example, we have observed differences across datasets in how late a peak can be and still be perceived as a 'L+H*' rather than a 'L*+ H', and it would be useful to know what stylistic or dialectal differences influence this category boundary. Analogously, the HiFo label was motivated by its usefulness for investigating the phonetics of pitch range variation as a cue to discourse topic structure or intentional structure (e.g. Grosz and Hirschberg 1992; Ayers 1994). This use of the HiFo label is related to a claim implicit in our decision not to mark break index values above the intonational phrase. There was a strong feeling among a sizeable group at the first workshop that the grouping of intonational phrases into 'utterances' and 'paragraphs' is qualitatively different from the grouping of morphemes into 'prosodic words'[3] dominated by intermediate phrases, and into intermediate phrases dominated by intonational phrases. The consensus understanding of the metrical hierarchy in phonological theory at the time

[3] The break index value 'o' was intended to mark a boundary between two orthographic words which is perceived to be considerably reduced in strength from a 'normal' word boundary. The MAE_ToB! conventions suggest that this sense of close grouping should be associated with such segmental sandhi phenomena as the flapping of final *ItI* in utterances such as *Got a dime?*, the palatalization of final *ItI* in *We sent you the cheque*, and so on-i.e., phenomena that have been cited by phonologists as evidence of multi-word prosodic constituents such as the 'Prosodic Word' or a 'Clitic Group' (see Hayes 1989; Selkirk 1996; Peperkamp 1999, and the references they cite for discussion of different theoretical views of these constituents). A break index value of '1' is then a

was that units such as prosodic word and intonational phrase constitute a non-recursive (and possibly even strictly-layered) tree (see, e.g. Nespor and Vogel 1986; Pierrehumbert and Beckman 1988). That is, a prosodic word cannot dominate other, smaller prosodic words at a lower level of grouping, and an intonational phrase cannot dominate other, smaller intonational phrases. Also, each such unit can be defined in terms of the distribution of categorical phonological markers in a way that makes the bracketing amenable to a finite numerical index. The phonological domain for the distribution of the phrase accent is a '3' because that is where the intermediate phrase is in the hierarchy; segmental sandhi phenomena such as Palatalization of a coda consonant by a following palatal (e.g. in *meet you)* does not occur across a domain bounded by a phrase accent, and a boundary tone never occurs within this domain. Units such as 'utterance' and 'paragraph', by contrast, seem to be recursive, reflecting something like a hierarchy of larger discourse topic and embedded smaller subtopics (see, e.g. Grosz and Sidner 1986). Work such as Lehiste (1975) further suggested that this recursive discourse structure will be reflected iconically in such continuous phonetic measures as the durations of inter-utterance pauses or the backdrop pitch range over successive intermediate phrases, rather than by direct categorical markers such as the distribution of boundary tones. Therefore, we decided not to label degrees of perceived disjuncture above the intonational phrase, so that the break indices could be interpreted as a direct implementation of the non-recursive prosodic hierarchy.[4] The MAE_ToB! conventions thus differ from other marking systems such as INTSINT, which propose symbolic labels for 'paragraph' and the like, based on the implicit claim of a single category of extra 'pitch resetting' at paragraph beginnings or a single discrete level of 'extra-low pitch' at paragraph ends (Hirst and Di Cristo 1999: 17).

Note that there is an ambiguity inherent in the conventions for relating the words and break indices to the audio recording. If only the end of each word is marked, as specified in the original MAE_ToB! conventions, then the Words tier labels a string of events. That is, each Words-tier tag (and the corresponding Break-Indices label) marks the boundary between a pair of words or between the last word and the following silence. As a result, the

'normal' word boundary. A more precise definition of these levels is desirable, but not yet feasible, because corpus research on such phenomena as flapping and palatalization lags considerably behind research on the phonetic correlates of prosodic grouping at the intermediate phrase and intonational phrase level.

[4] This meant omitting break indices 5 and 6 from the Price *et al.* (1991) model, since these two break index values could not be identified with a categorically marked level of prosodic structure such as the intonational phrase. Rather, they were intended to encode the percept of (possibly recursive) higher-level groupings above the intonational phrase.

onset of the first word and any mid-utterance pauses are not marked. However, there are many occasions, such as when training automatic speech recognition systems, when it is more useful to treat the 'words' as labelling intervals in the signal rather than single time points; for these purposes '<51L>' or '#' is often used to mark any word beginnings which are not coterminous with the end of the preceding word-i.e., at the beginning of the first word in the recording and after every pause. This was the expedient we adopted in converting the labels for the example utterances that accompany the *Guidelines to ToBI Labelling* into an EMU database.[5] These'<51L>' or '#' labels are thus the one exception to the rule that every word label must have a corresponding break index value. They are also a systematic source of inconsistency between 'dialects' of MAE_ToBI, although the inconsistency seems to be disappearing rapidly as more and more sites adopt this convention.[6]

This distinction between interval and event is encoded explicitly in the syntax for the Mise tier, which specifies two types of labels, differentiated by the interpretation of the time stamp. There are labels for localized events, such as 'disfl' (which marks the approximate time point where some disfluency is perceived), and there are paired labels for effects that occur over identifiable longer intervals, such as 'disfl <' and 'disfl>' (for the beginning and end of an identifiable stretch of disfluent speech). The dual syntax was our solution to the question of how to label a motley set of phenomena that could be either intervals (similar to words) or events (similar to breaks or tone targets). Another solution would have been to allow only the first type of label, but to permit the pairs of labels for more localized events to be separated by only one time frame, as in the EMU version of the *Guidelines to ToBI Labelling.*

We have illustrated the syntax of the Mise tier with the disfluency labels, because false starts and repairs are a perennial source of concern in analysing spontaneous speech. They are also a common source of 'ungrammaticality' or uncertainty about how to mark tones and break indices. One of the

[5] EMU is a set of tools for creating and analysing speech databases. It includes a powerful search engine that can find segments and events based on their sequential and hierarchical contexts. For example, if a MAE spoken language database has associated word labels, and if those labels are hierarchically organized into intermediate phrases and intonation phrases, with associated MAE_ToB! labels, it is straightforward to query for every instance in the database of a word with an associated L+H* pitch accent that is also the last accent in its intermediate phrase and followed by a !H- phrase accent. The EMU readable version of the *Guidelines to ToBI Labelling* is available at *http://www.shlrc.mq.edu.au/emu/emu-tobi.shtml.*

[6] We note that no site seems to have rigorously adopted the practice envisioned by the original ToBI group of marking silences automatically, on the Misc tier.

practical principles that we listed in Section 2.2 is that a ToB! annotation system needs to be reliable in order to be useful. This means that labels need to be applied consistently across sites and among transcribers at any given site, and that mechanisms must be provided for dealing with transcriber uncertainty and phonetic ambiguity. The MAE *ToBI Annotation Conventions* note that disfluencies 'are not automatically detectable, and the absence of markings for them makes it difficult to parse the Tones and Break-Indices tiers. For these reasons, transcribers are urged to mark disfluencies on the miscellaneous tier using "disfl<" and "disfl>" (or "disfl" if the disfluency is extremely localized)' (Beckman and Hirschberg 1994: 5). Related labels on the other tiers are the 'p' diacritic on the Break-Indices tier, to mark the 'perception of audible hesitation (for example, an abrupt cut-off or a prolongation)', and the '%r' label on the Tones tier, 'to indicate a "contour restart"-i.e., the initiation of a new intonational contour after a disruption'.

The MAE_ToB! conventions also specify a number of methods by which transcribers can indicate uncertainty in the absence of disfluency. For example, uncertainty about the strength of a break index is indicated by adding a '-' to the right of the index value. Thus, '4-' indicates that the transcriber found the preceding and following words to be somewhat more closely conjoined than is usual for words separated by a level 4 break index, but less clearly conjoined than those at a level 3 break index. Uncertainty on the Tones tier is indicated by a set of special symbols rather than by a diacritic, although all of these symbols include '?' as a final element. Thus, '*?' indicates uncertainty about whether or not a syllable has a pitch accent; '-?' indicates similar uncertainty about whether a phrase accent has occurred; and '%?' indicates uncertainty about whether a boundary tone has occurred. (Note that the latter two symbols should be accompanied by '3-' and '4-' on the Break-Indices tier.) Where tonal uncertainty concerns the type of tone, on the other hand, we employ 'X*?', 'X-?', and 'X%?' instead. So, while '*?' means 'I don't know whether this syllable is accented or not', 'X*?' means 'I believe that this syllable is accented but I don't know which pitch accent type to assign to it.' The first sort of uncertainty is exemplified in the first intonational phrase in Figure 2.1, where the labellers could not decide whether the first syllable of *okay* has a pre-nuclear L*, and in the last intonational phrase in Figure 2.3, where the labellers could not decide whether the apparent rise in Fo onto *Shore* represents a pre-nuclear H* or was just a point in the interpolation between the preceding L% boundary tone and the following H* on *Cab*. The latter sort of uncertainty is exemplified in the second intonational phrase in Figure 2.2, where the succession of downstepped accents on *southern* and *Iraqi* has compressed the pitch range to such an extent that there is no objective way to decide between a third !H* and a L* for the nuclear accent on *cities.*

Note that in many cases, labeller uncertainty can be attributed directly to aspects of the signal-i.e., to real phonological ambiguity that cannot be resolved just by training the transcriber to label more carefully. The MAE *ToBI Annotation Conventions* describe two such cases of phonological ambiguity which can be tagged with the Tones-tier uncertainty symbols: 'A typical case where "*?" might be used is for a very strong syllable in a part of an utterance between a prenuclear H* and a nuclear H*, where the Fo contour is flat and high because of the preceding and following tones, making it difficult to detect intervening H* accents. A typical case where "X*?" might be used is a part of an utterance where the labeller cannot tell whether an accent is a L* accent or a H* accent in a compressed pitch range.' The nuclear accent in the second intonational phrase in Figure 2.2 is an example of the latter sort of ambiguity. Thus, typical uses of '*?' and 'X*?' are cases of ambiguity involving mismatch between the perceived prominence of a syllable in the audio recording and the tonal markings of accentuation in the Fo record.

The '2' symbol on the Break-Indices tier was intended similarly to mark cases of ambiguity involving two types of mismatch between the perceived sense of disjuncture and the tonal markings of prosodic grouping. The symbol can mean that the perceived disjuncture is at break index level 3 or 4 (e.g. final lengthening and pausing that is appropriate for an intermediate phrase or intonational phrase), but that there is no clear indication of a phrase accent in the tone pattern. (The sense of pause between each pair of words in the sequence of downstepped accents in the second intonational phrase in Figure 2.2 is a good example.) Alternatively, the symbol 2 can mean that the perceived disjuncture is at break index level 1 (an ordinary phrase-internal juncture) despite the clear occurrence of a phrase accent or even a phrase accent and boundary tone sequence. (The steep fall from H* to L- after *Quincy* in Figure 2.3 is a good example.) Thus, break index '2' is not part of the metrical hierarchy *per se.* Rather, it provides a way to tag such cases of mismatch without jettisoning the definition of break indices 3 and 4 in terms of the otherwise well-governed coupling between tune and prosodic grouping. This provision distinguishes MAE_ToB! from earlier tagging systems for English which have no projection of metrical structure separate from the 'tone unit' (e.g. Crystal 1969) or which disclaim any correspondence between higher-level units in the metrical hierarchy and the domain(s) for associating tune to text (e.g. Gussenhoven 1990).

An aspect of this comparison to earlier tagging systems that was very salient in the discussion at the second workshop is the issue of redundancy and its effect on efficiency. That is, the explicit projection of the metrical structure onto a separate Break-Indices tier even though the higher level breaks are defined in terms of categorical tonal marks introduces redundancy that is somewhat at odds with the principle that ToB! conventions should be efficient. For example, every time a boundary tone is marked on the Tones tier, a 4 should be marked on the Break-Indices tier, and vice versa. The MAE *ToBI Annotation Conventions* acknowledge this redundancy and recommend that transcribers 'avail themselves of routines for automatically inserting redundant labels on either tier'. The function of the symbol 2 brings out the value of separating the function of the phrase accent in marking prosodic grouping from its function of autosegmental contrast. For example, the L- after *Quincy* in Figure 2.3 contrasts both with H - and with !H-, but there is no clear pause to separate *Quincy* (the caller's response to the operator's opening *What city please?)* from the following sentence (the caller's request for the telephone number). The redundancy allows the transcriber to modularize the tagging of the two separate structures and makes the tagging system theory-neutral by comparison to the implicit claim of strict correspondence in Crystal (1969) or the explicit claim of complete independence in Gussenhoven (1990). At the same time, the MAE_ToB! system does not exploit this modularity as gracefully as do several of the other ToB!-framework systems described in this book. The dual usage of the symbol '2' is one of the most awkward aspects of the MAE_ToB! conventions, and a common source of confusion for new transcribers. By contrast, the ToB! framework systems for Japanese, Korean, and Greek have introduced an explicit mismatch diacritic ('m') on the Break-Indices tier, so that break index 2 can be used for a well-defined level of the prosodic hierarchy for the language, such as the accentual phrase for Korean and Japanese, and the intermediate phrase for Greek (see Venditti 1997, this volume, Ch. 7; Jun this volume, Ch. 8; Arvaniti and Baltazani this volume, Ch. 4).[7] In the next section, we discuss some extensions of this modular design to other languages and to other phenomena.

2.5. EXTENSIONS OF TaBI

One of the first extensions of our work in developing MAE_ToB! was the application of the basic ToB! framework design to other languages, such as

[7] The intermediate phrase in Greek, like the intermediate phrase in English, is defined by the presence of a phrase accent after the nuclear pitch accent (see Grice, Ladd, and Arvaniti, 2000, for discussion of the cross-linguistic applicability of this concept). Thus, the use of 2 as a marker of two types of tones-breaks mismatch in English has resulted in different numbers corresponding to levels that are defined in the same way in the two languages.

northern German, Tokyo Japanese, and several other languages with less well-studied intonation systems. Although we made a firm decision at the first workshop that any tagging system we developed would have to be language-specific, we suspected that aspects of the MAE_ToBI design, particularly the explicit separation of autosegmental tonal content from hierarchical metrical structure, could be extended to other languages. Indeed, most of the systems described in this volume have implemented Tones and Break-Indices tiers, as well as some form of orthographic tier to provide the initial set of time stamps for relating tones and prosodic unit boundaries to the audio and Fo signals. As Jun (this volume, Ch. 16) points out, the applicability of the *Tones* versus *Break-Indices tier* structure to languages as typologically different as German, Japanese, and Cantonese suggests a prosodic universal: many languages seem to structure utterances into a hierarchy of prosodic units, at least some of which are categorically marked by the tone pattern.

It is important to emphasize, however, that the four labelling tiers described above, together with the audio and Fo records, are only the *obligatory* parts of the MAE_ToBI record. We had no expectation that these four tiers and two types of continuous record would suffice for tagging systems for all languages or even for all users of MAE databases. Rather, the originators of MAE_ToBI fully expected individual sites to add other ToBI tiers or other completely independent annotations as needed, in order to customize shared databases to their own purposes. For example, syntactic bracketing and part of speech can be projected in a hierarchical labelling system that is separate from the metrical hierarchy of break indices and the tonal projection in MAE_ToBI, in order to train algorithms for predicting intonational phrasing and accentuation in text-to-speech systems (see e.g. Hirschberg 1993; Hirschberg and Prieto 1994; Ostendorf and Veilleux 1994; Koehn *et al.* 2000; Hirschberg and Rambow 2001). Discourse structure also can be marked separately, and there are several standard discourse tagging systems available, based on different models of discourse structure. For example, the tagging system described in Nakatani *et al.* (1995) implements Grosz and Sidner's (1986) model of the speaker's hierarchy of discourse purposes as a basis for segmentation and global coherence, whereas Allen and Core's (1997) DAMSL system, Carletta *et al.'s* (1996) Map Task annotation scheme, and the annotation schema developed at the Discourse Resource Initiative workshops[8] implement more locally defined 'dialog act' models of discourse.

Like the MAE_ToBI system, a number of the discourse tagging schemes mentioned above have been tested in several inter-transcriber reliability

[8] See *http://www.georgetown.edu/luperfoy/Discourse-Treebankidri-ho*me.html.

exercises, and there has been much research in the ToBI community relating discourse structure tagged in these models to such continuous phonetic measures as syllable and pause durations, amplitude variation within and across discourse segment boundaries, pitch range relationships across successive intonational phrases, and so on. Studies now exist both for English (e.g. Grosz and Hirschberg 1992; Swerts and Ostendorf1997; Hirschberg and Nakatani 1998) and for several other languages for which ToBI framework models are available (see e.g. Venditti 2000 for Japanese). This research tends to support our prediction that discourse structure does not have a one-to-one correspondence to categorically marked prosodic domains above the intonational phrase. That is, there do not seem to be special tones or other categorical events that distinguish, say, the ends of paragraphs from the ends of sentences internal to a paragraph. Instead, the discourse hierarchy seems to be marked only by a fine, continuous control of variation in phonetic measures that can iconically reflect such (non-phonological) relationships as coordination versus embedding of discourse segment purposes.

Another common extension to the original MAE_ToBI tiers is to tag consonants and vowels on a separate autosegmental tier from the tonal projection. Any site which uses ToBI labelled data to train an automatic speech recognition or speech synthesis system does this, and the emerging convention is to call such a projection a *Phones* tier. At many such sites, a first-pass Phones-tier labelling is done automatically using an alignment programme, such as Aligner (Wightman and Talkin 1996) or some other similar HMM-based automatic transcription alignment system. For such sites, the Words-tier labels are then also derived automatically from the Phones alignment. A *Syllable* tier can also be added from the Phones tier, using a simple syllabification script, if desired. Stress labels can also be assigned, using an online dictionary in conjunction with the pitch accents on the Tones tier. Syrdal *et al.* (2001) describe such a syllable-tagging scheme, designed for training a variable-unit concatenative text-to-speech system. Ostendorf and Ross (1997) used similar tags in training an automatic intonation recognizer.

At the Fourth Prosodic Transcription Workshop, we discussed whether the Phones and Syllable tiers should be obligatory, but decided that it was not practical to make them so until alignment software becomes commonly and freely available. Some ToBI framework systems for other languages, however, have made other decisions. For example, the Pan-Mandarin ToBI system (M_ToBI, see Peng *et al.* this volume) specifies that a syllable-by-syllable segmental transcription of the utterance must be provided on a *Syllable* tier. Moreover, each interval marked on this M_ToBI tier must be labelled for its perceived degree of stress on an independent *Stress* tier. The contrast between stress levels 1 and 2 is defined by the categorical absence versus presence of an associated (lexical) tone, reminiscent of the definition of the levels 'unaccented' versus 'accented' in the English stress hierarchy. It is important to note that, while they are like the break indices in being a numerical index of perceived 'strength', these Stress tier labels differ from the break indices in marking intervals rather than events; they index the syllable's own strength rather than the following boundary strength. Thus, they constitute another metrical hierarchy that is independent of the metrical hierarchy of prosodic groups on the Break-Indices tier. That is, where the break indices correspond to a bracketing hierarchy (a metrical tree), the M_ToBI stress levels correspond to a rhythmic hierarchy (a metrical grid).

It would be easy to project the stress labels from the Syllable tier in Ostendorf and Ross (1997) and Syrdal *et al.* (2001) onto a similar independent Stress tier for MAE, although there are questions that need to be addressed before MAE_ToBI could make such a Stress tier obligatory. In particular, should prenuclear accent and nuclear accent project different levels of stress? And how many levels should there be below the accented/unaccented decision?

The labels for the (obligatory) Syllables tier in the Cantonese ToBI system (C_ToBI, Wong *et al.* this volume), by contrast, do not mark stress levels. Rather, this C_ToBI tier provides a transliteration in the roman alphabet of the standard Hong Kong reading of the Chinese characters and 'serves the function of the Words tier for sites that do not have a way to input and/or read Chinese characters' (Wong *et al.* this volume). Moreover, it is difficult to see how one could provide categorical definitions for consistently marking stress levels or syllables in Cantonese. Cantonese does have syllable-level lenition effects. However, these effects do not selectively target the segments in the less stressed syllable in a sequence of two, as in the superficially similar Mandarin lenition processes. Rather, the Cantonese fusion effects merely 'erase' the inter-syllable boundary, along a continuum from weakening or deleting the intervening consonant(s) to merging the two vowel qualities into an intermediate value. Moreover, it is typical for both syllables to maintain their status as tone-bearing units in the tonal projection even in cases where the consonant and vowel effects are so extreme that the syllable count is no longer clear in the segmental projection. Thus, Cantonese syllable lenition seems to be more a matter of prosodic grouping than of prominence-based rhythmic structure, and it is transcribed in C_ToBI by projecting a phonetic transcription of the affected syllable sequence onto a single prosodic unit on the *Foot* tier, while marking a value of 0 at the 'erased' boundary on the Break-Indices tier.

Comparing the (obligatory or imaginable) projections from the Syllables tier across the M_ToBI system, an extended MAE_ToBI system, and the C_ToBI system, then, we can make the following generalization about the usefulness of the sort of modularity that the ToBI framework promotes. Projecting numerical representations of the two different metrical hierarchies separately from each other as well as separately from the categorical tonal or segmental marks of any level of grouping or prominence brings out the fundamental similarity between (some varieties of) Mandarin and (some varieties of) English, while emphasizing the rather different role of the syllable in Cantonese. Cantonese lacks anything comparable to Mandarin or MAE syllable stress, and there is no basis for projecting a tier that encodes each syllable's metrical grid level. A similar conclusion also accords with research on Tokyo Japanese (e.g. Beckman 1986), on Mayali and many other Australian languages (e.g. Bishop and Fletcher this volume Ch. 12), on many dialects of Basque (Hualde *et al.* 2002), on Quebec French (e.g. Cedergren and Perrault 1994), and so on, and it is difficult to imagine a Stress tier for any of these language varieties. Thus, the rhythmic structuring of utterances by a hierarchy of syllable prominences seems to be not nearly so widespread as the structuring of utterances by a hierarchy of categorically marked prosodic units, and it is quite appropriate that the framework as a whole is called ToBI, and not, say, 'ToBISL' (see Peng *et al.* this volume Ch. 9).

A third very common extension ofMAE_ToBI is the recording of alternate analyses. For example, Syrdal *et al.* (2001) describe the use of a *Comments* tier to keep track of differences in proposed transcriptions when several transcribers are labelling a database together. (Figure 2.1 illustrates the use of the Misc tier for the same purpose.) The Syrdal *et al.* team found the Comments tier particularly useful in the initial stages of labelling a new voice or a new speaker style. Periodic discussion of recurring patterns in the sets of alternative transcriptions allowed the transcribers to compare across utterances to calibrate their criteria for distinguishing between superficially similar contours such as two rising shapes, and to articulate cues such as subtle differences in timing, slope, or transition extent. In this way, annotating many utterances together brought out important questions for future research, such as the possibility of inter-speaker or perhaps inter-dialectal differences in the location of the boundary between the L+H* and L*+H categories along a peak timing continuum. This question could be investigated in several ways. For example, one might elicit imitative productions using a synthetic continuum, as in Pierrehumbert and Steele (1989). Alternatively, one might gather and label a suitably large multi-speaker corpus recorded in dialogue tasks that are designed to elicit tokens of these two accent types.

The *Phonetic* tier in the expansion of the Korean ToB! system proposed by Jun (this volume, Ch. 8) is similarly motivated. That is, the use of a separate tier to tag 'non-canonical' tone targets at the edges of accentual phrases seems difficult to justify if the term 'phonetic' is taken too literally-i.e., if the labels on this tier are viewed as a substitute for a more direct phonetic representation such as the Fo contour. However, Jun's Phonetic tier seems intended simply as an interim device to keep track of different kinds of apparent mismatch between the Tones and Break-Indices tiers, in a system that was perhaps codified too early, on the basis of an overly restricted set of speech styles. If larger and larger numbers of completely fluent L-ending accentual phrases are discovered as the system is applied to a richer variety of spontaneous speech styles produced by Seoul speakers who do not command any other variety of the language, then the Phonetic tier labels could become the basis for a rigorous discussion of how to revise the K_ToB! annotation conventions to provide better coverage. Alternatively, given the recent history of rural in-migration in South Korea (more than 30 per cent of the population now lives in the Seoul area), it may turn out that L-ending accentual phrases signal solidarity with some other major regional variety, such as Chonnam or Kyungsang Korean. In that case, the 'non-canonical' tone targets might be better accommodated, not by expanding the inventory of accentual-phrase tones posited for the standard Korean model, but by setting up the appropriate alternative inventories for the other dialects involved, and adding a *Code* tier to indicate points of code-switching between the standard and regional dialect inventories, as proposed for different varieties of Mandarin by Peng *et al.* (this volume).

In the same way, we can imagine something like Syrdal and colleagues' Comments tier or Jun's Phonetic tier becoming an extremely useful tool in the initial stages of investigating whether the MAE_ToB! system can be applied to other varieties of English, as in Fletcher and Harrington's (1996) study of rise times in standard Anglo-Australian English and Fletcher and Warren's (2000) investigation of the different nuclear rises in both standard Australasian varieties (see also Section 15.2 in Fletcher, Grabe and Warren, this volume). That is, for example, we currently do not know whether the greater prevalence of rising boundary configurations in Australian English compared to MAE and SBE is due to a different pragmatics for such sequences as H+!H* !H H% or to a slightly different inventory of pitch accent types. Until we have enough data to suggest a definitive phonological analysis, a more cautious procedure might be to note examples of the potentially categorically different nuclear rise types on a separate Phonetic tier. However, we would caution against thinking of such a tier as a 'truly phonetic' representation, since this invites a misinterpretation of the status of symbolic tags in the ToB! framework that the originators of the MAE_ToB! system hoped to preclude. Since the original ToB! has been criticized as 'somewhat vague' (Nolan and Grabe 1997: 259) regarding the status of its symbolic tags, we would like to clarify our conception of their function here.

2.6. THE PHONETIC REPRESENTATIONS IN MAE_ToB!

As Pierrehumbert puts it in a recent paper, the original MAE_ToB! system 'is at the level of abstraction of a broad phonemic transcription, or rationalized spelling system, such as those of Korean and Finnish. Just as a broad phonemic transcription for any language must be guided by the phoneme inventory of that language (as revealed by the lexical contrasts), a ToB!-style transcription of the prosody and intonation of any language must be guided by an inventory of its prosodic and intonation patterns' (Pierrehumbert 2000: 26). This point cannot be emphasized too strongly. Symbols imply symbolic categories, and their use implies that the labeller has recourse to extensive prior research of the sort that would allow positing an inventory of categories relevant for the language variety being transcribed. A theory-neutral or 'non-linguistic' symbolic transcription is impossible, even in theory. This is true even for transcribing consonants and vowels, whether in nonsense words in the labeller's own language or in utterances of a language that the labeller does not speak. Using the IPA or any similar alphabetic device to label speech data brings with it two strong theoretical claims: that utterances in the language being transcribed can be segmented into consonant and vowel categories, and that the language has consonant and vowel inventories that will not be too different in design from the inventories of already analysed languages on which the IPA is based. The fact that the IPA has been used successfully by field researchers to build phonemic analyses of hundreds of languages justifies the assumption of these claims as a plausible first hypothesis, but it does not change their status as theoretical claims. Using the IPA does not prevent arguments about the 'phonetic' data in cases where one or the other claim is not fully justified, as can be appreciated by reading, for example, Coleman (1996) versus Dell and Elmedlaoui (1996) regarding the segmental status of Berber stop releases.

This characterization of symbolic transcription as a necessarily phonological act holds even more strongly for tonal categories, since even varieties of the same language can differ markedly in their bases for segmenting the tone contour. For example, in a language with lexical tone, the pitch fall in a disyllabic word specified for a high tone on the initial syllable followed by low tone on the second in one variety might sound like a falling lexical tone in another variety where tone contrasts are specified over the whole word. Or, when the word is uttered in isolation, it might sound exactly like a disyllabic phrase with high lexical tone on the initial syllable followed by a tonally unspecified second syllable before a low boundary tone. Similarly, in a language where pitch accents are a relevant descriptive notion, a rising accent in phrase-final position in one variety might sound like a rise from a low pitch accent to a high boundary tone in another variety. Or it might be confused with a complex boundary tone in a third variety where the initial target is not anchored to any syllable because of influence from a substrate language that does not have accent. Transcribing the 'same' HL or LH pattern for all of the types in each set of cases cannot elucidate the differences among the varieties. There is no substitute for a detailed comparison of the Fo contours in a controlled examination of the 'same' tone sequences in other phrasal contexts and other pragmatic conditions.

A related point is that the fundamental frequency contour is an obligatory phonetic representation in the ToB! framework. Psychoacoustics research might eventually yield a better phonetic representation of the pitch contour, but a continuous phonetic representation can never be replaced by even the most detailed symbolic encoding of pitch events. To put it another way, symbolic tone labels in the ToB! framework are intended to 'tag' the intonation contour and not to 'encode' it. A tag is a pointer for retrieving phonologically relevant portions of the fundamental frequency and audio signals. It is not a symbolic representation of purportedly language-neutral pitch levels (Chao 1920) or pitch movements (Hirst, Di Cristo, Le Besnerais, Najim, Nicolas, and Romeas 1993). This is an especially useful way to understand the relationship between a MAE_ToB! transcription and the signal, because of the long history of research on the inventory of contrasts in varieties to which MAE_ToB! is known to be applicable. Someone who subscribes to a slightly different model of MAE or SBE intonation (such as the model of Crystal 1969 or Gussenhoven 1984) and who wants to investigate the phonetics of a category that is explicit in the alternative model but not in MAE_ToB! should still be able to use a database that has been labelled with MAE_ToB! tags, because there probably will be a correspondence between the tags that MAE_ToB! provides and the tags that the researcher would have used if labelling in the alternative model. Although it is not likely to be a simple one-to-one correspondence (cf. Roach 1994), this correspondence probably will be lawful and hence useful. For example, anyone who wants to look at the factors determining the slope or extent of the pitch movements within a 'sliding head' (i.e., a pre-nuclear stretch with a peaky alternation of rise and fall instead of the gently downstepping trend that was the most common pattern for the head in Crystal's 1969 SBE corpus) can search for MAE_ToBI sequences such as H* L+H*, L+H* L+H*, and so on, to identify relevant instances of pre-nuclear pitch falls in a MAE_ToBI-labelled database. After listening to the utterance transcribed with the MAE_ToBI Tones-tier labels shown in (1),[9] the researcher who subscribes to Crystal's (1969) analysis, is free to replace the MAE_ToBI tones transcription in (1) with the tonetic stress marks in (2). Note that while these two transcriptions differ from each other in the phonological analysis assumed, with (1) describing the tune as a sequence of rising accents and (2) describing it as a sequence of falling accents, both differ even more fundamentally from Hirst's (1999: 73) INTSINT 'narrow phonetic' transcription of the sliding head in (3), which only notes the alternation of rise and fall without analysing any part of the tune as the contrastive pitch event that is distinctively anchored to the accented syllable.

(1) There's a lovely yellowish old one.
        L+H* L+!H* L+!H* L-L%

(2) There's a`lovely`yellowish`old one.

(3) There's a LOVEly YELlowish OLD one.
    [⇑ ↓ ↑ ↓   ↑ ⇓   ]

The status of the symbolic tags in a ToBI framework system thus differs qualitatively from the status of the labels in the INTSINT system (Hirst *et al.* 1993; Hirst and Di Christo 1999), as well as from the labels on the 'pitch movement' tier in IViE (Grabe, Nolan, and Farrar 1998; Grabe 2000). The formulators of INTSINT wanted a tool for making the 'equivalent of a narrow phonetic transcription' (Hirst and Di Christo 1999: 14), something that could be used to record 'pitch points' or 'targets' in intonation contours of a language even before the relevant categories for anchoring tune to text can be known. Thus, the INSTINT symbols ⇓ and ↓ are used to transcribe a fall in pitch, whether it is the necessarily steep interpolation from the peak of a L+H* rising accent to an immediately following L-phrase accent in the English utterance in (3), or the variably steep interpolation from the peak of a

[9] See http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html for this utterance. Hirst (1999: 73) reports that the sliding head 'has been described as typical of Scottish accents' and suggests that it 'is probably gaining ground throughout England possibly due to the influence of American speech where the pattern is very common'. Our impression is that it is more characteristic of Australian and New Zealand varieties, particularly those with a strong Scottish English substrate, than it is of mainstream American varieties—see, e.g. Fletcher and Harrington 1996; Ainsworth 2000.

LH accent at the beginning of one word to the valley of the LH accent at the beginning the next word in the Finnish example in (4). (The latter is Hirst and Di Cristo's interpretation of an untranscribed Finnish example in Iivonen 1999. See Välimaa-Blum 1988 for our analysis of the fall as an interpolation between word-initial LH pitch accents, which will be more or less steep depending on the lengths of the words.)

(4) LAIna LAInaa LAInalle LAInan.
[⇑ ↓ ↑ ↓ ↑ ↓ ↑ ⇓]
'Laina lends Laina a loan.' (= Example (6) in Hirst and Di Cristo 1999: 16).

The IViE 'pitch movement' labels similarly were formulated to be an 'auditory phonetic transcription' of pitch movements which the labeller uses to 'capture the phonetic realization of a pitch accent in Fo, at least as far as that is possible with a set of discrete tone labels' (Grabe 2000). *The IViE Labelling Guide* justifies this on pedagogical grounds (new labellers tend to 'rely too heavily on Fo' and do not develop skills for 'careful listening'), and on the grounds that the discrete symbols make for easier compilation and comparison of phonetic realization patterns ('H*+L, for instance, can be realized as hM-l, as mH-l, as hH-l, as mH-l (peak lag) or lH (truncation)'). In other words, like the INTSINT arrows and braces in (3) and (4), the IViE 'pitch movement' labels are 'phonetic' because they attempt to encode a downsampled Fo contour.

Calling a symbolic tier in the ToBI framework a 'Phonetic tier', by contrast, signifies something quite different. Rather than an attempt at a language-neutral encoding of the Fo pattern, a Phonetic tier label in the ToBI framework is a variety-specific tag, a way to keep track of possible phonological analyses of tune and of tune-text association at a stage when a research team has looked at enough data to make plausible guesses but does not yet have enough knowledge to specify a definitive inventory of contrasting intonations and prosodic patterns. The labels are discrete not because they are a grossly downsampled encoding of the Fo contour, but because they are hypotheses about discrete phonological categories. They still bear the same relationship to continuous phonetic representations that the final set of tags on the Tones tier will bear. That is, they will not replace the audio and Fo signals, but merely provide a convenient way to retrieve all instances where a particular hypothesis was made (along with instances of relevant potentially contrasting categories) so as to make possible a subsequent more penetrating examination of the phonetic record and of the associated phonological, syntactic, and pragmatic contexts. For this reason, our own advice for how best to fill the needs that *The IViE Labelling Guide* identifies is to use the Fo contour to the full, at least until a better representation of pitch is developed. The pedagogical ends, which are important, can be satisfied by exercises designed to teach new transcribers to effectively 'listen' to the Fo contour (e.g. McGory 2000). The more seasoned researcher, similarly, cannot hope to avoid the labour of finding effective ways to control the materials, of devising good Fo and other phonetic measures, and of creating new experimental tasks to test competing analyses, as in Liberman and Pierrehumbert (1984), Pierrehumbert and Steele (1989), Silverman and Pierrehumbert (1990), Hirschberg and Ward (1991), Shattuck-Hufnagel, Ostendorf, and Ross (1994), and a host of other studies of tone alignment, tone scaling, and other aspects of MAE intonational categories.

Another way to understand the difference between these two approaches is to consider how vowel systems might be compared across two varieties of the same language. The formulators of IViE opt for a narrow symbolic transcription: 'We can describe all varieties of English as having three (historically) short front vowels i, e, and a. But if we want to describe the difference between the English of New Zealand and Yorkshire we need the phonetic categories [ɪ e ɛ] and [ɨ ɛ a̠]' (Nolan and Grabe 1997: 260). In the view that guided the development of the MAE_ToBI standard, a different approach would be adopted. Decades of research in sociolinguistics suggests that a more illuminating way to capture the difference in timbre between the corresponding New Zealand and Yorkshire vowel categories is to represent the vowels in the two dialects in terms of the distribution of formant values in appropriately large and varied databases (cf. Hindle 1979; Labov 1994; Watson *et al.* 1998; Docherty and Foulkes 1999). The question then reduces to the mechanics of how to tag the databases. Vowel formant values can be extracted from a database however the vowels are tagged, so long as they are tagged consistently within each database. That is, it does not matter whether the tags are [ɪ], [e], and [ɛ] for New Zealand versus [ɨ], [ɛ], and [a̠] for Yorkshire, or 'i', 'e' and 'a' (or 'I', 'E' and '@', or 'ih', 'eh', and 'ae') for both. What matters is that all of the instances of any one of these three historically short front vowels be tagged in the same way within a given database and differentiated from the historically long vowel counterparts, so that researchers can retrieve all instances of the different categories and compare their formant values between the two varieties.

Of course, this characterization of the ToBI approach begs the question of how one decides that two vowels belong to the same or to different categories within any one variety, or how vowel categories might correspond across two different varieties. In the comparison between New Zealand and Yorkshire, we know the correspondences because these three vowels are phonemes which contrast a large number of words that are common to both dialects, and because a substantial body of philological research shows that (especially by comparison to historically back vowels) these short front vowel phonemes developed in a very uniform way across the lexicon in the dialects of English that were married in making the New Zealand variety. Lexical tone categories and their correspondences in cognate words across varieties of Mandarin are somewhat messier, but still fairly easy to establish by comparison to the tones of English, where the categories typically do not contrast sets of morphemes but constitute pragmatic morphemes in their own right (see, e.g. Ladd 1980; Gussenhoven 1984; Ward and Hirschberg 1985; Pierrehumbert and Hirschberg 1990). However, a large body of research using a variety of techniques has given us a good idea of what the tonal categories are for SBE and MAE, and there is a large body of research on intonation systems of other languages that can be tapped as a source of ideas about what questions to ask in deciding what the categories are for an unstudied (or under-studied) language variety.

In looking at a relatively unstudied variety of English, for example, an obvious first set of questions to ask would be: 'What is the history of this variety? Was there contact with a substrate language that might lead us to expect it to lack such MAE and SBE categories as stress and nuclear pitch accent? If so, how can we characterise patterns that seem to correspond to these categories in MAE and SBE?' These are especially valid questions to ask for varieties such as Hawaiian English, Singapore English, or West African English (see, e.g. Vanderslice and Pearson 1967; Lim 1997; Gut 2000); it would be a mistake to begin one's analysis of these varieties by having MAE or SBE speakers tag the syllables they perceive to be stressed. On the other hand, if the understudied variety clearly is related to MAE and SBE with respect to the applicability of the notions of stress and accent, then one can proceed differently. The researcher who is a speaker of SBE or MAE might enlist a native speaker of the other variety to collaborate in asking questions such as the following. 'Do we consistently hear nuclear accent as falling in the same place in utterances elicited in the same controlled contexts in both of our varieties? If so, what is the Fo pattern around the nuclear-accented syllable in a broad focus SVO declarative utterance in the other variety? Is the Fo pattern (and the syntax) the same in a context that puts narrow focus on the object NP? What about on the subject NP, or the verb? Can we elicit subject narrow focus in longer declarative sentences, to see what happens when there are few versus many words following the nuclear accent? What is the Fo pattern around the nuclear accented syllable in the inverted broad focus yes-no question that is the counterpart to the short SVO statement and which does not presuppose its answer? If the Fo contour in this case is a rise from the nuclear syllable, do we see the same rising pattern in a yes-no question when the speaker expects a "yes" answer, or does the rise begin at a higher level? What happens if the yes-no question focuses on the subject NP or on the verb?'

Some of these questions can be addressed also by looking at a suitable corpus that one eventually intends to tag in a ToBI framework system, particularly if one can collaborate with a native speaker consultant, as in Daly and Warren (2001). Developing elicitation protocols for recording suitable corpora, such as the Map task (Anderson *et al.* 1991), is thus another important research endeavour in its own right. Whatever the materials, however, the hard slogging work of addressing such questions cannot be circumvented by trying to force the analysis of tunes in the other variety into a transcription system designed for SBE and MAE. Researchers approaching a previously undescribed variety should not try to use the original MAE_ToBI system (or Crystal's system, or Halliday's system, or any other transcription system designed for SBE or MAE) as if it were a variety-neutral phonetic transcription system for intonation and prosody. Couching the research programme within a transcription framework that has been used to describe these better-studied varieties can help formulate relevant questions for the initial analysis and for later comparison across the varieties, but it cannot do more than that. A transcription system and the framework for developing it are two separate things.

In sum, applying the ToBI framework to develop a transcription system for a new variety presumes that the development team has access to an established body of research specific to the variety for which the transcription system is intended. If there is no such body of research, the development team must do the necessary research. Knowing a great deal about several other varieties of the same language, and being able to state that knowledge in a common framework, can help to establish relevant controls from early in the research endeavour. However, there is no shortcut around actually doing the research. A ToBI framework system devised for one language variety cannot be assumed to be applicable even to other varieties of the same language, without first establishing appropriate intonational and prosodic analyses for each variety. Any claim that the symbolic tags are comparable across varieties must be based on a thorough variety-specific analysis of each of the varieties involved, as in Fletcher and Warren's (2000) study of Fo contours for high rising terminals in both Australasian varieties of English. This is so because the Tones-tier labels in a ToBI framework system are comparable to a broad phonemic representation of consonants and vowels, and not to a narrow phonetic one.

2.7. THE WORK AHEAD FOR MAE_ToBI

That said, we can think of several useful ends that a more 'allophonic' transcription might serve. One is to record features of productions in regions where dialect contact has established a 'mixed code' in which the different distributions of phonetic values for a common phonological category come to have a kind of distinctive status as badges of contrasting social affiliation. As more and more utterances of Mainstream American English are transcribed with MAE_ToBI, it is likely that we will discover systematic differences among different communities. We would be surprised if no one finds mixed codes incorporating, say, features of the African-American Vernacular English intonation system into a predominately MAE utterance for stylistic effect (cf. Hay *et al.* 1999). It is our hope that a *Code* tier (as proposed in the Pan-Mandarin ToBI system—see Peng *et al.* this volume) will provide the right approach for capturing such phenomena, but further experience with the MAE_ToBI system in a variety of contexts will be necessary to test its appropriateness.

A number of other issues might be illuminated by an allophonic transcription of consonant and vowel segments, of the sort that Veillieux and Shattuck-Hufnagel (1998) and Jurafsky *et al.* (2002) are already using to study the prosody of function words. The question of whether to add a Stress tier to MAE_ToBI (as in the proposed Pan-Mandarin system) probably is one that will require a closer study of the consonant and vowel qualities that are associated with stressed syllables at different levels of the prominence hierarchy. Some instances of low inter-transcriber agreement regarding the presence or absence of a pitch accent suggest that speakers can use segmental rhythms to create the sense of accentual prominence in the absence of any tonal markings of pitch accent. Some instances of break index 2 (as in the utterance in Figure 2.2) similarly suggest that speakers can use the rhythms of intermediate phrasing to set off words as focally prominent without ending the current phrase by pronouncing a phrase accent. In general, MAE_ToBI is vague about the segmental effects relevant for differentiating break index levels when there are no tonal marks. Wightman *et al.* (1992) and others have examined durational correlates of juncture, particularly at the higher break index levels. Pierrehumbert and Talkin (1992) and Dilley *et al.* (1995), similarly, have shown that vowels and sonorant consonants tend to have more or less glottalized variants when they occur at the beginning of an intermediate phrase or an intonational phrase, particularly if the initial syllable also is accented. However, tagging of break index level 0 (e.g. prosodic-word internal, as in *gimme, doncha*) versus break index level 1 (normal word boundary, as in *give me, don't you*) tends to be less consistent across teams of experienced MAE_ToBI transcribers, because we do not have a good understanding of the phonetic bases for break index 0 in cases where the segmental correlates are less obvious than in the almost lexicalized examples given here (see Syrdal *et al.* 2001). A good study of where tap variants of /t/ and /d/ occur in MAE and the Australasian varieties would help clarify whether the MAE *ToBI Annotation Conventions* were too sanguine in citing this as a straightforward clue to the creation of a constituent smaller than a 'normal' word in these varieties.

Of course, these issues will not be addressed adequately just by adding a more allophonic tagging of consonants and vowels in the relevant cases; research aimed at finding a quantitative phonetic representation also is needed. The Fo contour is relatively easy to calculate as a phonetic representation of pitch, but other phonetic properties, such as degree of glottalization (or creak) and degree of breathiness are not as well studied. The field would be well served by the development of phonetic representations of these voice qualities that can be applied to continuous speech, and not just to sustained vowels. We are encouraged about this direction of research by the recent increase in attention to the problem, with closer interaction among researchers who study voice quality from a variety of viewpoints, including its role in phonological contrast, its systematic variation with prosodic structure, and its range and mechanisms of variation in pathology (cf. Shattuck-Hufnagel *et al.* 2001). Other important efforts involve the attempt to develop classification systems for laryngealization events in continuous speech (Batliner *et al.* 1993), and to characterize their distribution (Kohler 1994; Hagen 1997). Work on accent in Dutch (e.g. Sluijter and van Heuven 1996) suggests to us that phonetic measures of voice quality could illuminate some of the cases of perceived phrase-level prominence in the absence of pitch accent. Such measures clearly are needed also to represent some intonational contrasts, such as the two different interpretations (incredulity versus uncertainty) of the rise-fall-rise tune (Hirschberg and Ward 1992). Current acoustic phonetic representations of timing (e.g. Campbell and Isard 1991) and 'rhythmicity' (e.g. Ramus *et al.* 1999; Grabe and Low, 2002) also seem quite crude by comparison to our understanding of articulatory dynamics (e.g. Browman and Goldstein 1990; Beckman and Cohen 2000; Munhall *et al.* 2000). Basic research to devise acoustic measures of timing phenomena that are as good as our phonetic representation of pitch would be useful. For example, better phonetic representations of timing and rhythmicity should illuminate our understanding of effects such as the 'chanted' or 'stylized' variant of H* !H- L%, which constitutes the 'calling contour' (Ladd 1980). Better measures of timing and of voice quality also might increase inter-transcriber reliability for break index 3 versus break index 4 in the tonally ambiguous cases of L- versus L-L% and H- vs. H-L%.

As the above suggests, we think that comparing points of greater and lesser inter-labeller reliability is useful for suggesting avenues of necessary further research. This means that the ToBI endeavour could also benefit from more work on inter-transcriber consistency, and from the development of other metrics for establishing correspondences between the criteria that different transcribers use to distinguish categories that they perceive as more or less similar (see McGory *et al.* 1999). In the same vein, it would be useful to establish the correspondences between MAE_ToBI transcriptions and transcriptions made using other prosody annotation schema. Comparing points of high 'inter-system reliability' with points where correspondences are more difficult to establish could illuminate areas where the underlying models of intonational phonology need work. For example, the IViE 'Phonological tier' labels, like Gussenhoven's (1984) model of SBE and all of the traditional British systems on which it is based, differs from MAE_ToBI in allowing no leading tones, such as the leading L tone that distinguishes MAE_ToBI L+H* from H*, or the leading H tone that distinguishes MAE_ToBI H+!H* (= Pierrehumbert's H+L*) from !H* or L*. Since IViE is based on British systems that are designed for SBE, we might ask whether this difference between the transcription systems can be attributed to a more substantive difference between the two dialects. If this is the case, Grice (1995) shows that inter-dialect differences cannot be the sole explanation, since SBE also clearly has an accent category with a leading tone falling onto a lower tone (which Grice transcribes as a separate H+L* pitch accent type). The MAE_ToBI distinction between H* and L+H*, on the other hand, is more controversial. Work by Ladd and colleagues (Ladd 1993; Ladd and Morton 1997) suggests that, in SBE at least, the H* versus L+H* contrast might be a gradient difference in prominence associated with 'normal' versus 'expanded' pitch ranges rather than a strictly categorical binary distinction. The contrast also causes more inter-transcriber disagreement in transcribing MAE than any other accent pair (see Silverman *et al.* 1992; Pitrelli *et al.* 1994; Syrdal *et al.* 2001). In these two cases of accents with leading tones, comparing across the MAE_ToBI and IViE transcription systems augments the comparison across transcribers to highlight places where both systems might be revised.

Establishing correspondences across transcription schemes also has a more practical utility. Implementing known correspondences into automatic translation algorithms would expand the repertoire of databases available to all researchers, whatever their theoretical orientations.

In short, as members of the original development team, we would characterize the original ToBI system as we would characterize the ToBI framework as a whole: it is an ongoing research programme rather than a set of 'rules' cast in stone for all time. Despite its 'unfinished' nature, however, the effort that went into its development clearly has been worthwhile. In addition to what MAE_ToBI has taught us about the process of transcription system development, a number of results from using the system signal its value. These results include the overall high level of transcriber agreement, the system's productiveness in encouraging and guiding the development of similar systems for other languages, and its usefulness as a communal corpus creation tool even in its current state. There are now many MAE_ToBI corpora, and a good number of these are publicly available. For example, a large part of the BU FM Radio news database has been annotated using MAE_ToBI by teams at Boston University and MIT, and these tags have been used in a large variety of research projects ranging from the training of an intonation recognizer and an intonation synthesis system (Ostendorf and Ross 1997; Ostendorf and Ross 1999) to the definitive study of stress shift as early accent placement (Shattuck-Hufnagel *et al.* 1994). Large parts of the native speaker Map Task dialogues in the Australian National Database of Spoken Language (ANDOSL) have been annotated with MAE_ToBI labels by teams at Macquarie University and University of Melbourne, and this database also has been used for a variety of purposes such as training duration rules for an Australian English TTS system (Fletcher and McVeigh 1993) and identifying spontaneous speech materials for use in psycholinguistics experiments testing the role of accent in pronoun resolution (Stirling *et al.* 2001). A large portion of the MAGIC corpus at Columbia University has been MAE_ToBI labelled and used for exploring the relationship of various syntactic, semantic, and discourse features to intonational features such as accent and phrasing, including a test of Bolinger's idea of semantic weight as a determinant of accentuation (Pan and McKeown 1999).

Thus, the vision which inspired Victor Zue to convene the initial workshop, and the willingness of the various sets of participants to persevere through negotiations whose intensity would make Kofi Annan shudder, has resulted in the creation of large labelled databases, a better understanding of the strengths and weaknesses of the underlying phonological theories, the development of ToBI-like systems for other languages, and even in the development of contrasting systems inspired by the concreteness and explicit assumptions of the ToBI approach. We look forward to further developments in our understanding of spoken prosody, such as the ToBI-framework systems described in this book.

REFERENCES

AINSWORTH, H. (2000), 'Telling Tales in Taranaki: Evidence of Regional Variation in New Zealand English', paper presented at the workshop on Varieties of English Intonation and Prosody, Victoria University of Wellington, 12–14 Dec.

ALLEN, J. and CORE, M. (1997), *DAMSL: Dialog Act Markup in Several Layers.* Online MS available at http://www.cs.rochester.edu/research/trains/annotation/RevisedManual.

ANDERSON, A. H., BADER, M., BARD, E. G., BOYLE, E., DOHERTY, G., GARROD, ISARD, S., KOWTKO, J., MCALLISTER, J., MILLER, J., SOTILLO, C., THOMPSON, H. S., and WEINERT, R. (1991), 'The HCRC Map Task Corpus', *Language and Speech*, 34: 351–66.

ANDERSON, S. R. (1978), 'Tone Features', in V. A. Fromkin (ed.), *Tone: A Linguistic Survey* (New York: Academic Press), 133–76.

ARMSTRONG, L. E. and WARD, I. C. (1926), *A Handbook of English Intonation* (Cambridge, UK: Heffer and Sons).

ARVANITI, A. and BALTAZANI, M. (this volume Ch. 4), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora'.

AYERS, G. M. (1994), 'Discourse Functions of Pitch Range in Spontaneous and Read Speech', *Ohio State University Working Papers in Linguistics*, 44: 1–49.

BATLINER, A., BURGER, S., JOHNE, B., and KIESSLING, A. (1993), 'MUESLI: A Classification Scheme for Laryngealizations', in D. House and P. Touati (eds.), *Proceedings of an ESCA Workshop on Prosody, Lund 1993*, 176–9.

BECKMAN, M. E. (1986), *Stress and Non-Stress Accent* (Dordrecht: Foris).

—— and AYERS, G. M. (1994), *Guidelines for ToBI Labelling.* Online MS and accompanying files available at http://www.ling.ohio-state.edu/~tobi/ame_tobi.

—— and COHEN, K. B. (2000), 'Modeling the Articulatory Dynamics of Two Levels of Stress Contrast', in M. Horne (ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce* (Dordrecht: Kluwer), 169–200.

—— and HIRSCHBERG, J. (1994), *The ToBI Annotation Conventions.* Online MS. Available at http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.

—— and PIERREHUMBERT, J. B. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–309.

BOLINGER, D. L. (1951), 'Intonation: Levels versus Configurations', *Word*, 7: 199–210.

—— (1958), 'A Theory of Pitch Accent in English', *Word*, 14: 109–49.

Bolinger, D. L. (1964), 'Around the Edge of Language: Intonation', *Harvard Educational Review*, 34: 282–93.

Browman, C. P. and Goldstein, L. (1990), 'Tiers in Articulatory Phonology, With Some Implications for Casual Speech', in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I* (Cambridge: Cambridge University Press), 341–76.

Bruce, G. (1977), *Swedish Word Accents in Sentence Perspective* (Lund: Lund University).

—— (1982), 'Developing the Swedish Intonation Model. *Working Papers in Phonetics* (Department of Linguistics and Phonetics, University of Lund), 22: 51–116.

Campbell, W. N. and Isard, S. D. (1991), 'Segment Durations in a Syllable Frame', *Journal of Phonetics*, 19: 37–48.

Carletta, J., Isard, A., Isard, S., Kowtko, J., and Doherty-Sneddon, G. (1996), *HCRC Dialogue Structure Coding Manual. Technical Report HCRC/TR*-82. Available online at http://www.hcrc.ed.ac.uk/publications/tr-82.ps.gz.

Cedergren, H. J. and Perrault, H. (1994), 'Speech Rate and Syllable Timing in Spontaneous Speech', *Proceedings of the 1994 International Conference on Spoken Language Processing* (Yokohama, Japan), 1087–90.

Chao, Y. R. (1920), 'A System of Tone Letters', *Le maître phonétique*, 45: 24–7.

Charniak, E. (2000), 'A Maximum-Entropy-Inspired Parser', *Proceedings of the ANLP-NAACL 2000* (Seattle, Washington), 132–9.

Coleman, J. (1996), 'Declarative Syllabification in Tashlhit Berber', in J. Durand and B. Laks (eds.), *Current Trends in Phonology: Models and Methods* (Salford: European Studies Research Institute, University of Salford), Vol. 1, 177–218.

Collins, M. (1999), 'Head-Driven Statistical Models for Natural Language Parsing', Ph.D. dissertation (University of Pennsylvania).

Cooper, W. E. (1976), 'Syntactic Control of Timing in Speech Production: A Study of Complement Clauses', *Journal of Phonetics*, 4: 151–71.

Crystal, D. (1969), *Prosodic Systems and Intonation in English* (Cambridge, UK: Cambridge University Press).

Daly, N. and Warren, P. (2001), 'Pitching it Differently in New Zealand English: Speaker Sex and Intonation Patterns', *Journal of Sociolinguistics*, 5: 85–96.

Dell, F. and Elmedlaoui, M. (1996), 'Nonsyllabic Transitional Vocoid in Imdlawn Tashlhiyt Berber', in J. Durand and B. Laks (eds.), *Current Trends in Phonology: Models and Methods* (Salford: European Studies Research Institute, University of Salford), Vol. 1, 217–44.

Dilley, L., Shattuck-Hufnagel, S., and Ostendorf, M. (1995), 'Individual Differences in the Glottalization of Vowel-initial Syllables', *Journal of the Acoustical Society of America*, 97: 3418–19.

Docherty, G. J. and Foulkes, P. (1999), 'Instrumental Phonetics and Phonological Variation: Case Studies from Newcastle upon Tyne and Derby', in P. Foulkes and G. J. Docherty (eds.), *Urban Voices: Accent Studies in the British Isles* (London: Arnold), 47–71.

FLETCHER, J. and BISHOP, J. (this volume Ch. 12), 'Intonation in Six Dialects of Bininj Gun-Wok'.

—— GRABE, E. and WARREN, P. (this volume Ch. 14), 'Intonational Variation in Four Dialects of English: The High Rising Tune'.

—— and HARRINGTON, J. (1996), 'Timing of Intonational Events in Australian English', *Proceedings of the Sixth Australian International Conference on Speech Science and Speech Technology* (Adelaide), 611–15.

—— and MCVEIGH, A. (1993), 'Syllable and Segment Duration in Australian English', *Speech Communication*, 13: 355–65.

—— and WARREN, P. (2000), 'Variation in Rises and Rises in Varieties', paper presented at the workshop on Varieties of English Intonation and Prosody, Victoria University of Wellington, 12–14 Dec.

GEE, J. P. and GROSJEAN, F. (1983), 'Performance Structures: A Psycholinguistic and Linguistic Appraisal', *Cognitive Psychology*, 14: 411–58.

GRABE, E. (2000), *The IViE Labelling Guide, Version 2.* Online MS available at http://www.mml.cam.ac.uk/ling/ivyweb/guide.html.

—— and LOW, E. L. (2002), 'Durational Variability in Speech and the Rhythm Class Hypothesis', in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7* (Berlin: Mouton de Gruyter), 515–43.

—— NOLAN, F., and FARRAR, K. (1998), 'IViE—a Comparative Transcription System for Intonational Variation in English', *Proceedings of the 1998 International Conference on Spoken Language Processing* (Sydney) (distributed on CD-Rom by the Australian Speech Science and Technology Association).

GRICE, M. (1995), 'Leading Tones and Downstep in English', *Phonology*, 12: 183–233.

—— REYELT, M., BENZMÜLLER, R., MAYER, J., and BATLINER, A. (1996), 'Consistency in Transcription and Labelling of German Intonation with GToBI', *Proceedings of the 1996 International Conference on Spoken Language Processing* (New Castle, DE: Citation Delaware), 1716–19.

—— LADD, D. R., and ARVANITI, A. (2000), 'On the Place of Phrase Accents in Intonational Phonology', *Phonology*, 17: 145–87.

GROSZ, B. and HIRSCHBERG, J. (1992), 'Some Intonational Characteristics of Discourse Structure', *Proceedings of the 1992 International Conference on Spoken Language Processing* (Banff: University of Alberta), 429–32.

—— and SIDNER, C. (1986), 'Attention, Intentions, and the Structure of Discourse', *Computational Linguistics*, 12: 175–204.

GUSSENHOVEN, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

—— (1990), 'Tonal Association Domains and the Prosodic Hierarchy in English', in S. Ramsaran (ed.), *Studies in the Pronunciation of English* (London: Routledge), 27–37.

GUT, U. (2000), Session on West African Englishes. Workshop on Varieties of English Intonation and Prosody, Victoria University of Wellington, 12–14 Dec.

HAGEN, A. (1997), 'Linguistic Functions of Glottalizations and their Language Specific Use in English and German', thesis, Erlangen University.

Halliday, M. A. K. (1967), *Intonation and Grammar in British English* (The Hague: Mouton).

Hay, J., Jannedy, S., and Mendoza-Denton, N. (1999), 'Oprah and /ay/: Lexical Frequency, Referee Design and Style', *Proceedings of the 14th International Congress of Phonetic Sciences* (San Francisco, CA) (distributed on CD-Rom by the Regents of the University of California).

Hayes, B. (1989), 'The Prosodic Hierarchy in Meter', in P. Kiparsky and G. Youmans (eds.), *Perspectives on Meter* (New York: Academic Press), 203–60.

Hindle, D. M. (1979), 'The Social and Situational Conditioning of Phonetic Variation', Ph.D. dissertation (University of Pennsylvania).

Hirschberg, J. (1993), 'Pitch Accent in Context: Predicting Intonational Prominence from Text', *Artificial Intelligence*, 63: 305–40.

—— and Nakatani, C. (1998), 'Acoustic Indicators of Topic Segmentation', in *Proceedings of the 1998 International Conference on Spoken Language Processing* (Sydney) (distributed on CD-Rom by the Australian Speech Science and Technology Association).

—— and Prieto, P. (1994), 'Training Intonational Phrasing Rules Automatically for English and Spanish Text-to-Speech', in *Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis* (New Paltz, NY), 159–62.

—— and Rambow, O. (2001), 'Learning Prosodic Features Using a Tree Representation', *Proceedings of Eurospeech 2001* (Alborg: Center for Personkommunication).

—— and Ward, G. (1992), 'The Influence of Pitch Range, Duration, Amplitude, and Spectral Features on the Interpretation of L*+H L H%', *Journal of Phonetics*, 20: 241–51.

Hirst, D. (1999), 'Intonation in British English', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems* (Cambridge, UK: Cambridge University Press), 56–77.

—— and Di Cristo, A. (1999), 'A Survey of Intonation Systems', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems* (Cambridge, UK: Cambridge University Press), 1–44.

——, —— Le Besnerais, M., Najim, Z., Nicolas, P., and Roméas, P. (1993), 'Multi-Lingual Modelling of Intonation Patterns, *Proceedings of the ESCA Workshop on Prosody (Lund Working Papers in Linguistics and Phonetics, 41)*, 204–07.

Hualde, J. I., Elordieta, G., Gaminde, I., and Smiljanić, R. (2002). 'From Pitch Accent to Stress Accent in Basque', in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7* (Berlin: Mouton de Gruyter), 547–84.

Iivonen, A. (1999), 'Intonation in Finnish', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems* (Cambridge, UK: Cambridge University Press), 311–27.

Jun, S.-A. (this volume Ch. 8), 'Korean Intonational Phonology and Prosodic Transcription'.

—— (this volume Ch. 16), 'Prosodic Typology'.

Jurafsky, D., Bell, A., and Girand, C. (2002), 'Phonological Variation as Evidence for Lexical Representation of Homonyms', in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7* (Berlin: Mouton de Gruyter), 3–34.

Kingdon, R. (1939), 'Tonetic Stress Markers for English', *Maître Phonétique*, 54: 60–4.

Koehn, P., Abney, S., Hirschberg, J., and Collins, M. (2000), 'Improving Intonational Phrasing with Syntactic Information', *Proceedings of ICASSP 2000* (Istanbul).

Kohler, K. J. (1994), 'Glottal Stops and Glottalization in German', *Phonetica*, 51: 38–51.

Labov, W. (1994), *Principles of Linguistic Change* (Cambridge, MA: Blackwell).

Ladd, D. R. (1980), *The Structure of Intonational Meaning: Evidence from English* (Bloomington, IN: Indiana University Press).

—— (1983), 'Phonological Features of Intonational Peaks', *Language*, 59: 721–59.

—— (1990), 'The Metrical Representation of Pitch Register', in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I* (Cambridge, UK: Cambridge University Press), 35–57.

—— (1993), 'Constraints on the Gradient Variability of Pitch Range (or) Pitch Level 4 Lives!' in P. Keating (ed.), *Papers in Laboratory Phonology III* (Cambridge, UK: Cambridge University Press), 43–63.

—— (1996), *Intonational Phonology* (Cambridge, UK: Cambridge University Press).

—— and Morton, R. (1997), 'The Perception of Intonational Emphasis: Continuous or Categorical?', *Journal of Phonetics*, 25: 313–42.

Leben, W. (1973), 'Suprasegmental Phonology', Ph.D. dissertation (Massachusetts Institute of Technology).

Lehiste, I. (1960), 'An Acoustic-Phonetic Study of Internal Open Juncture', *Phonetica*, suppl.

—— (1975), 'The Phonetic Structure of Paragraphs', in A. Cohen and S. Nooteboom (eds.), *Structure and Process in Speech Perception* (Berlin: Springer-Verlag), 195–206.

Liberman, M. and Pierrehumbert, J. (1984), 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aronoff and R. Oehrle (eds.), *Language Sound Structure* (Cambridge, MA: MIT Press), 157–233.

Lim, L. (1997), 'Intonation Patterns Characterising Three Ethnic Varieties of English in Singapore: Observations and Implications', in *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications* (Athens, Greece), 207–10.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994), 'The Lexical Nature of Syntactic Ambiguity Resolution', *Psychological Review*, 101: 676–703.

Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1993), 'Building a Large Annotated Corpus of English: The Penn Treebank', *Computational Linguistics*, 19: 313–30.

Mayo, C., Aylett, M., and Ladd, D. R. (1997), 'Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI', in *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications* (Athens, Greece), 231–4.

McCawley, J. D. (1968), *The Phonological Component of a Grammar of Japanese* (The Hague: Mouton).

McGory, J. T. (2000), 'Linguistics 795T: Practicum in English Intonation', course materials (Ohio State University).

——, Herman, R., and Syrdal, A. (1999), 'Using Tone Similarity Judgments in Tests of Intertranscriber Reliability', *Journal of the Acoustical Society of America*, 106: 2242.

Munhall, K. G., Kawato, M., and Vatikiotis-Bateson, E. (2000), 'Coarticulation and Physical Models of Speech Production', in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V* (Cambridge, UK: Cambridge University Press), 9–39.

Nakatani, C. H., Groz, B. J., Ahn, D. D., and Hirschberg, J. (1995), *Instructions for Annotating Discourse.* (Center for Research in Computing Technology, Technical Report Number TR-21-95; Cambridge, MA), available online at: ftp://ftp.pitt.edu/dept/lrdc/edtech/jmoore/emd/nakatani-etal-guide.ps.Z.

Nakatani, L., O'Conner, K., and Aston, C. (1981), 'Prosodic Aspects of American English Speech Rhythms', *Phonetica*, 38: 84–106.

Nespor, M. and Vogel, I. (1986), *Prosodic Phonology* (Dordrecht: Foris).

Nolan, F. and Grabe, E. (1997), 'Can "ToBI" Transcribe Intonational Variation in British English?', in *Proceedings of the ESCA Workshop on Intonation: Theory, Models and Applications* (Athens, Greece), 259–62.

O'Malley, M. M., Kloker, D., and Dara-Abrams, B. (1973), 'Recovering Parentheses from Spoken Algebraic Expressions', *IEEE Transactions in Audio and Electroacoustics*, AU-21: 217–20.

Ostendorf, M. and Ross, K. (1997), 'A Multi-Level Model for Recognition of Intonation Labels', in Y. Sagisaka, N. Campbell, and N. Higuchi (eds.), *Computing Prosody* (New York: Springer-Verlag), 291–308.

—— and Ross, K. (1999), 'A Dynamical System Model for Generating Fundamental Frequency for Speech Synthesis', *IEEE Transactions on Speech and Audio Processing*, 7: 295–309.

—— and Veilleux, N. (1994), 'A Hierarchical Stochastic Model for Automatic Prediction of Prosodic Boundary Location', *Computational Linguistics*, 20: 27–54.

Palmer, H. E. (1922), *English Intonation with Systematic Exercises* (Cambridge, UK: Heffer).

Pan, S. and McKeown, K. (1999), 'Word Informativeness and Automatic Pitch Accent Modeling', in *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* (University of Maryland, College Park, MD).

Peng, S., Chan, M. K.-M., Tseng, C., Huang, T., Lee, O., and Beckman, M. E. (this volume Ch. 9), 'Towards a Pan-Mandarin System for Prosodic Transcription'.

Peperkamp, S. (1999), 'Prosodic Words', *GLOT International*, 4(4): 15–16.

Pierrehumbert, J. B. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

—— (2000), 'Tonal Elements and their Alignment', in M. Horne (ed.), *Prosody: Theory and Experiment. Studies Presented to Gösta Bruce* (Dordrecht: Kluwer), 11–36.

——and Beckman, M. E. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

——and Hirschberg, J. (1990), 'The Meaning of Intonation Contours in the Interpretation of Discourse', in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271–311.

——and Steele, S. (1989), 'Categories of Tonal Alignment in English', *Phonetica*, 46: 181–96.

——and Talkin, D. (1992), 'Lenition of /h/ and Glottal Stop', in G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II* (Cambridge, UK: Cambridge University Press), 90–116.

Pitrelli, J. F., Beckman, M. E., and Hirschberg, J. (1994), 'Evaluation of Prosodic Transcription Labelling Reliability in the ToBI Framework', in *Proceedings of the 1994 International Conference on Spoken Language Processing* (Yokohama, Japan), 123–6.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991), 'The Use of Prosody in Syntactic Disambiguation', *Journal of the Acoustic Society of America*, 90: 2956–70.

Ramus, F., Nespor, M., and Mehler, J. (1999), 'Correlates of Linguistic Rhythm in the Speech Signal', *Cognition*, 73: 265–92.

Roach, P. (1994), 'Conversion between Prosodic Transcription Systems: "Standard British" and "ToBI" ', *Speech Communication*, 15: 91–9.

Selkirk, E. O. (1978), *On Prosodic Structure and its Relation to Syntactic Structure* (Bloomington, IN: Indiana University Linguistics Club).

——(1995), 'Sentence Prosody: Intonation, Stress, and Phrasing', in J. Goldsmith (ed.), *The Handbook of Phonological Theory* (Cambridge, MA: Blackwell), 550–69.

——(1996), 'The Prosodic Structure of Function Words', in J. Morgan and K. Demath (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Mahwah, NJ: Lawrence Erlbaum Associates), 187–213.

Shattuck-Hufnagel, S., Gerratt, B. R., and Kreiman, J. (eds.) (2001), Special issue of the *Journal of Phonetics* on voice quality, 29(4): 363–482.

——Ostendorf, M. and Ross, K. (1994), 'Stress Shift and Early Pitch Accent Placement in Lexical Items in American English', *Journal of Phonetics*, 22: 357–88.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992), 'TOBI: a Standard for Labeling English Prosody', in *Proceedings of the 1992 International Conference on Spoken Language Processing* (Banff, Canada), 867–70.

——and Pierrehumbert, J. (1990), 'The Timing of Prenuclear High Accents in English', in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology I* (Cambridge, UK: Cambridge University Press), 72–106.

Sluijter, A. M. C. and van Heuven, V. J. (1996), 'Spectral Balance as an Acoustic Correlate of Linguistic Stress', *Journal of the Acoustical Society of America*, 100: 2471–85.

Srinivas, B. and Joshi, A. K. (1999), 'Supertagging: An Approach to Almost Parsing', *Computational Linguistics*, 25: 237–65.

Stirling, L., Fletcher, J., Mushin, I., and Wales, R. (2001), 'Representational Issues in Annotation. Using the Australian Map Task Corpus to Relate Prosody and Discourse Structure', *Speech Communication*, 33: 113–34.

Swerts, M. and Ostendorf, M. (1997), 'Prosodic and Lexical Indications of Discourse Structure in Human-Machine Interactions', *Speech Communication*, 22: 25–41.

Syrdal, A. K., Hirschberg, J., McGory, J. T., and Beckman, M. (2001), 'Automatic ToBI Prediction and Alignment to Speed Manual Labeling of Prosody', *Speech Communication*, 33: 135–51.

Trager, G. L. and Smith, D. L. (1951), *An Outline of English Structure* (Norman, OK: Battenburg Press).

Trim, J. K. M. (1959), 'Minor and Major Tone Groups in English', *Le maître phonétique*, 112: 26–9.

Välimaa-Blum, R. M. (1988), 'Finnish Existential Clauses—Their Syntax, Pragmatics and Intonation', Ph.D. dissertation (Ohio State University).

Vanderslice, R. and Pearson, L. S. (1967), 'Prosodic Features of Hawaiian English', *Quarterly Journal of Speech*, 53: 156–66.

Veilleux, N. and Shattuck-Hufnagel, S. (1998), 'Phonetic Modification of the Syllable /tu/ in Two Spontaneous American English Dialogues', *Proceedings of the 1998 International Conference on Spoken Language Processing* (Sydney) (distributed on CD-Rom by the Australian Speech Science and Technology Association), 2527–30.

Venditti, J. J. (1997), 'Japanese ToBI Labelling Guidelines', *Ohio State University Working Papers in Linguistics*, 50: 62–72.

—— (2000), 'Discourse Structure and Attentional Salience Effects in Japanese Intonation', Ph.D. dissertation (Ohio State University).

—— (this volume Ch. 7), 'The J_ToBI model of Japanese intonation'.

Ward, G. and Hirschberg, J. (1985), 'Implicating Uncertainty: The Pragmatics of Fall-Rise Intonation', *Language*, 51: 747–76.

Watson, C. I., Harrington, J., and Evans, Z. (1998), 'An Acoustic Comparison between New Zealand and Australian English Vowels', *Australian Journal of Linguistics*, 18: 185–207.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., and Price, P. J. (1992), 'Segmental Durations in the Vicinity of Prosodic Phrase Boundaries', *Journal of the Acoustical Society of America*, 91: 1707–17.

—— and Talkin, D. T. (1996), 'The Aligner: Text-to-Speech Alignment using Markov Models', in J. van Santen, R. Sproat, J. Olive, and J. Hirschberg (eds.), *Progress in Speech Synthesis* (New York: Springer-Verlag), 313–23.

Wong, P. W.-Y., Chan, M. K.-M., and Beckman, M. E. (this volume Ch. 10), 'An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese'.

3

German Intonation in Autosegmental-Metrical Phonology

Martine Grice, Stefan Baumann, and Ralf Benzmüller

3.1. INTRODUCTION

English, Dutch, and German are often claimed to have very similar prosody and intonation, an observation that might be related to the fact that they all belong to the West Germanic language family. All three have a stress-timed rhythm with left-headed feet, and they all make use of a number of different pitch accents for highlighting information, and of edge tones for delimiting phrases. In broad focus contexts, they all place the nuclear pitch accent on the final argument in the intonation phrase.

Within the autosegmental-metrical framework there are essentially two major approaches to German intonation. On the one hand, there are accounts such as those by Féry (1993) and Grabe (1998), which follow Gussenhoven's (1984) analysis of Dutch. On the other there is GToBI, a consensus system developed by Martine Grice, Matthias Reyelt, Ralf Benzmüller, Anton Batliner, and Jörg Mayer (Grice *et al.* 1996; Reyelt *et al.* 1996), which is closely related to the original English ToBI (Mainstream American English ToBI, henceforth MAE_ToBI: Beckman and Hirschberg 1994; Beckman and Ayers-Elam 1997; Beckman *et al.* this volume Ch. 2) and the analysis of English by Pierrehumbert (1980) upon which MAE_ToBI is based.

The major differences between these approaches are twofold. First, in Gussenhoven's model pitch accents, like feet, are always left-headed. This means that a given pitch accent cannot account for the pitch before the

We would like to thank Caroline Féry, Jürgen Trouvain, and Jörg Peters for insightful comments on earlier drafts.

accented syllable to which it is associated. GToBI and MAE_ToBI, by contrast, have not only left-headed but also right-headed pitch accents. The latter account both for the pitch on the accented syllable and for the pitch immediately before it, in which case the tone to the left is referred to as a leading tone. Second, Gussenhoven analyses nuclear contours as a combination of a pitch accent and an intonation phrase boundary tone. GToBI and MAE_ToBI instead postulate an additional tone after the final pitch accent in the phrase. This extra tone is referred to as the phrase accent.

At first glance, it might appear that the differences result from the fact that one group of researchers consider German to be like English, whereas the other take it to be like Dutch. However, since Gussenhoven also disputes the existence of leading tones and phrase accents in English, it becomes obvious that the differences are of a theoretical rather than a typological nature.

In this chapter we shall begin by looking at the relatively theory-neutral traditional literature on German intonation, often based on auditory impressions with a great deal of phonetic detail. We then go on to give an overview of the autosegmental-metrical literature which builds on Gussenhoven's work, and provide a detailed exposition of GToBI. Finally, we offer motivation for the analysis used in GToBI as compared to the other autosegmental-metrical models.

3.2. ACCOUNTS OF GERMAN INTONATION

Traditionally, intonation has been analysed either in terms of tonal configurations, i.e. pitch contours whose direction is important, or in terms of levels, where the pitch range is divided up into a number of discrete levels. In the latter, intonation contours are derived from sequences of these levels. We shall first examine the configurations-based accounts and then go on to look at early levels approaches, and at the more recent levels-based autosegmental-metrical accounts.

3.2.1. *Configurations-based approaches*

Early accounts of German intonation, such as those by von Essen (1964), Pheby (1975), Kohler (1977), and Fox (1984) are mainly auditory-based and didactically oriented, representing intonation patterns with a detailed interlinear transcription of the pitch of each syllable of an utterance. All of the above are akin to British-style analyses, e.g. Crystal (1969) and Halliday (1967), in treating intonation in terms of dynamic pitch contours, and in attributing particular importance to the nucleus (Halliday's 'tonic' and Pheby's 'Tonstelle') which is said to be the utterance's most prominent syllable. For Pheby and

Fox, the contour by which tunes are classified starts at the nuclear syllable and continues to the end of the phrase. In the British School, this contour is referred to as the nuclear tone.

A somewhat different configurations-based approach has been proposed by Selting (1995). Her aim is to develop a descriptive system for the analysis of spontaneous dialogues. Selting's model is auditory-based, partly supported by instrumental analysis. She distances herself from the British School in regarding intonation contours as holistic units with no special status assigned to the nucleus. For Selting, unlike for the early auditory studies of German, intonation does not reflect grammatical structure, but is used for signalling 'prosodically cohesive' units relevant for discourse organization, e.g. in the construction of turns. Selting's is what can be termed an 'overlay approach' (Ladd 1996). She not only specifies the shape of individual pitch accents but also whether the upper and lower limits of the pitch range are globally declining, inclining, or level.

Kohler treats pitch accents as pitch peaks which may be aligned in different ways with the text. In his more recent instrumentally based work (e.g. 1991), he shows that it is possible to differentiate between three types of peak (early, medial, and late) and to assign pragmatic interpretations to them: An early peak, where the peak is on the prenuclear syllable, marks a self-evident or established fact. A medial one, where the peak is around the middle of the accented syllable, indicates a new fact. A late peak, occurring towards the end of the accented syllable or even on the following syllable, places emphasis on a new fact and/or represents greater involvement on the part of the speaker than is the case with a medial peak (1991: 160).

3.2.2. *Early levels-based approaches*

Not all accounts of German have been configurations-based. In the early 1960s, Moulton (1962), in the American structuralist tradition of Pike (1945) and Trager and Smith (1951) described the intonation of German in terms of distinct pitch levels (Moulton had three rather than the usual four pitch levels; he did not discuss level 4, the emphatic level). Like Trager and Smith, Moulton also had what is called 'terminal contours', indicating whether the pitch was rising, falling, or sustained at the very end of the phrase, thus making the approach more a mixture of levels and contours than a strict levels approach. The first strict levels approach, Isačenko and Schädlich (1966), reduced the number of levels to two. They resynthesized utterances on high and low monotone pitch levels with a step up or down from one level to another, either before the accented syllable (which they refer to preictic, the ictus being the accented syllable) or after it (postictic). The preictic fall is equivalent to Kohler's early peak contour, the postictic fall to the medial or late peaks. We shall return to early peak contours in the discussion of GToBI pitch accents.

3.2.3. *Autosegmental-metrical accounts*

More recent levels-based approaches have been developed within the autosegmental-metrical framework. Autosegmental-metrical (AM) is a term coined by Ladd (1996) to refer to the approaches to intonation which developed following on from the seminal work of Pierrehumbert (1980). These approaches generally make use of minimally two (H and L), maximally three (H, L, and M) levels for the description of intonation. These may have a prominence-lending function, being grouped together into pitch accents. Pitch accents are generally either monotonal (e.g. H*) or bitonal (e.g. L*+H). The starred tone is said to phonetically align with the accented syllable, although recent research has shown that alignment is more complicated than this (Arvaniti *et al.* 1999). If the unstarred tone precedes the starred tone, it is referred to as a leading tone. If it follows the starred tone, it is a trailing tone. As we have already seen, one of the differences between Pierrehumbert's model and GToBI on the one hand and the rest of the AM models for German on the other is that the former have both leading and trailing tones, whereas the latter have only trailing tones. We deal with this issue in more depth in Section 3.4.1.

Tones may also have a delimitative function, acting as initial or final edge tones of intonationally relevant phrases. In the models surveyed below, tones are phonetically realized as coordinates on the frequency-time axis. However, the scaling of these tones when they are combined into accent or accent-edge tone clusters is not always transparent. As we shall see, the use of H and L tones differs considerably among the different accounts. Furthermore, scaling is affected by downstep, which lowers the pitch of certain H tones, or even upstep which raises the pitch of both H and L tones.

The AM models of German intonation include those of Wunderlich (1988), Uhmann (1991), Féry (1993) and Grabe (1998), the latter two in turn influenced by Gussenhoven's (e.g. 1984) account of the intonational systems of English and Dutch. Since GToBI is based on autosegmental-metrical theory, we shall briefly survey each approach,[1] list their inventories of pitch accents

[1] From the two models based on Gussenhoven's system, we investigate in detail here the earlier of the two (i.e. Féry 1993).

and boundary tones, and examine how they describe commonly occurring tunes, to pave the way for the comparison with GToBI in Section 3.4.

We begin by looking at the work of Wunderlich (1988). Like most other German phonologists, he emphasizes the 'grammaticalized' functions of intonation, especially sentence modality and focus-background structure (cf. e.g. Altmann *et al.* (1989), who provide extensive experimental data relating to such functions). Wunderlich's inventory of intonation patterns consists of single accents, accent-accent sequences and accent-boundary tone sequences. They are listed below, along with the contexts in which they occur, if any are given:

H*	peak accent (Gipfelakzent)—default accent
H* H L*	bridge accent (Brückenakzent)—multiple foci, contrast
%H L*	falling-to-low accent (Fallend-Tiefakzent)—exclamations
L* H%	low accent-to-rise (Tiefakzent-Steigend)
L* H (H%)	echo accent (Echoakzent)—echo questions
H* H	left bridge pillar (linker Brückenpfeiler)—beginnings of lists

Wunderlich claims that each pattern may have several functions, paying particular attention to the functions of the bridge accent. This accent type is said to be typical of German sentences with multiple foci: after the peak accent (H*) the pitch stays high until it sharply falls on or to the accented syllable (L*). A floating H tone accounts for the high pitch between the two accents. It surfaces as what appears to be at once a trailing H for the first and a leading H for the second accent. %H L*, where %H stands for a phrase initial boundary tone, is used in the second half of a bridge accent if there is a sentence-medial phrase boundary separating the two halves of the bridge. GToBI would transcribe the final accent in the bridge accent as having a leading H tone, which would correspond to both the floating H in H* H L* and the %H in %H L*.

Uhmann (1991) provides a book-length and therefore more detailed account of German intonation. Like Wunderlich, she restricts her investigation to the prosodic marking of grammatical functions, especially the relation between intonation and focus-background structure. Uhmann's inventory consists of an optional initial boundary tone L% or H% (the default being mid), an obligatory final boundary tone L% or H%, and four pitch accents L*, H*, L*+H, H*+L. The nuclear accent is always bitonal; prenuclear accents can be mono- or bitonal. Tonal targets before the accented syllables (i.e. leading tones) are not considered necessary.

Uhmann treats boundary tones as the phonological correlates of phrasing, and pitch accents as the phonological correlates of the focus feature. She assigns more or less distinctive meanings to the pitch accents: prenuclear L*+H functions as topic marker, L* highlights background constituents, H* highlights focus or background constituents, and H*+L represents the default focus accent.

She lists the following patterns and the sentence modality with which they co-occur:

H*+L L%	declaratives, w-questions
L*+H H%	echo questions, yes/no-questions
H*+L H%	yes/no-questions (nuclear pitch accent marked)
L*+H L%	w-questions (nuclear pitch accent marked)

The account proposed by Féry (1993) also deals with the influence of focus structure on German tonal patterns. Her inventory slightly deviates from Uhmann's: She has the same four bitonal pitch accents. The monotonal ones are derived by phonological rule rather than being underlyingly part of the inventory. Féry does not posit an initial boundary tone, although she does have an optional high (H%) terminal boundary tone which is used when the pitch at the boundary is considerably higher than the pitch of the trailing tone of the nuclear accent. Following Gussenhoven (1984), she claims that all pitch accents are left-headed and underlyingly bitonal (H*+L or L*+H),[2] although her inventory has exceptions which will be discussed below. In a sequence of pitch accents (e.g. H*+L H*+L), a prenuclear accent can be linked to a nuclear accent, either partially, in which case the trailing tone of the prenuclear accent is associated with a syllable near the nuclear accent resulting in what appears to be a surface tritonal accent with a leading L tone (e.g. H* L+H*+L), or totally, in which case the trailing tone of the prenuclear accent is deleted, resulting in H* H*+L.

One aim of Féry's study is to give a comprehensive overview of the variety of tonal patterns occurring in German. She describes in detail two different types of hat pattern (Wunderlich's bridge accent). The first is exemplified in (1).

(1) H* H*+L
BALD ist sie DA
SOON is she THERE
'SOON she'll be THERE' (Féry 1993: 150)

[2] For clarity of exposition, a '+' sign is transcribed between tones belonging to a pitch accent. This notational convention taken from Pierrehumbert (1980) is not used by Féry.

It is usually realized in one phrase. The second type of hat pattern surfaces as a sequence of a rise and a fall (L*+H H*+L), an example of which is given in (2).

(2) L*+H H*+L
geSCHLAfen hat KEIner von uns
SLEPT had NONE of us
'NONE of us had SLEPT' (Féry 1993: 129)

Féry claims that this pattern is obligatorily realized in two intermediate phrases, and that it can be distinguished from the first type of hat pattern by the contexts in which it occurs. She claims that the first type has a very wide usage, whereas the second is restricted to topic-comment sentences, i.e. it consists of a topic accent (L*+H) and a focus accent (H*+L), or to several kinds of contrast, gapping, or clefts.

Féry postulates six different nuclear contours, giving the contexts in which they typically occur, or the meanings which they impart with respect to sentence modality, pragmatic interpretation and speaking style.

H*+L	declaratives, w-questions, wishes, imperatives
L*+H	progredient intonation, questions (e.g. echo questions), uncertainty/indignation
H*+L H%	questions, threats
L*+H+L	implying 'of course', slightly menacing
H+H*+L	TV-reporter style
H*+M	calling contour

However, the structure L*+H+L is problematic, since Féry has to take recourse to a bitonal trailing tone. Such a tone is not needed anywhere else in the system and one might consider whether the contour would have been better described as a L*+H followed by L%. She rejects this analysis, presumably because of the way the tones are distributed across the syllables; in terms of phonetic alignment, the rise-fall is not simply a mirror image of the fall-rise (which she represents as H*+L H%). GToBI, which, as we shall see, has not only trailing tones but also two different boundary tones, is able to capture a rising-falling pattern as L*+H L-(%), where H is a trailing tone. This is not parallel to the falling-rising pattern H* L-H%, where the alignment of the fall is more variable, depending on the position of postnuclear stresses and is represented as an L intermediate phrase boundary tone (also referred to as a phrase accent, see Sections 3.3.3 and 3.4.2 below). GToBI therefore correctly predicts that the alignment of tones with the text is different in each pattern. Furthermore, the self-evident or 'of course' meaning assigned to L*+H+L, which Féry also refers to as a late peak contour ('später Gipfel'; 1993:

96), is not the same as the meaning of Kohler's late peak. In fact, as we have seen in 3.2.1, Kohler assigns 'self-evidence' to a completely different contour: the early peak.

The description of stylized contours, e.g. calls, poses another problem for Féry's model. She represents the calling contour as H*+M, thus adding a mid pitch level to her inventory. Although there are arguments for treating calling contours as distinct from other intonational phenomena since they are often chanted, and might therefore necessitate a more musical notation, GToBI accounts for calling contours using the regular intonational inventory: the accented syllable is high, H*, and the step down represented as a H- phrase accent is downstepped (!H-).

Although Féry explicitly claims that her pitch accents are left-headed, that is, the starred tone is always on the left, she allows for an accent where it is evident that the pitch before the starred tone is relevant for the accent shape. This is the early peak accent H+H*+L. We shall see in Section 3.4.1 that this contour is also represented in GToBI, albeit with a different sequence of tones.

3.3. GToBI

3.3.1. *Preliminaries*

GToBI is a set of conventions for labelling German intonation with the aim of being easy to learn, reliable, and adaptable for different labelling purposes. It is therefore not a strictly phonological description of German intonation. It was developed between 1995 and 1996 by researchers from Saarbrücken, Stuttgart, Munich, and Braunschweig with a view to facilitating the exchange of prosodically annotated data. A cross labeller consistency test with the consensus system was carried out and reported on in Grice *et al.* (1996) and Reyelt *et al.* (1996). Results showed that labellers were able to use GToBI consistently, most of them having learned it within a short period of time from printed training materials and accompanying sound files (Benzmüller and Grice 1997) and with little or no individual coaching. The GToBI training materials have been updated and are available via the GToBI home page http://www.coli.uni-sb.de/phonetik/projects/Tobi/gtobi.html. What is presented in this chapter is a slightly modified version of the original GToBI.

Part of the flexibility of GToBI is achieved by different levels of description—so called tiers. The consensus system comprises at least three tiers, containing labels for words, tones, and break indices. As a general principle, information is only encoded if it cannot be derived (automatically) from labels from other tiers or from the speech signal. Thus, only mismatches, or non-default correspondences are transcribed by hand.

The words tier provides an orthographic transcription of the words spoken. On the tones tier the perceived pitch contour is transcribed in terms of pitch accents and boundary tones, with symbols for pitch range modifiers such as downstep and upstep placed immediately before the affected tone. Phrase boundary strength information is recorded in the break index tier. Other information may be added in an optional miscellaneous tier. An example screen shot with speech waveform, labels for tones, words, break indices, and miscellaneous information, along with the Fo contour is given in Figure 3.1. The labels will be discussed below.

Much of the information is only interesting if it involves relating the tiers to each other. For instance, in the tone tier there is information as to which pitch accents are realized. Now, one of the functions of pitch accents is to highlight particular words. Information as to which words are highlighted by means of a pitch accent can only be gleaned from relating the position in time

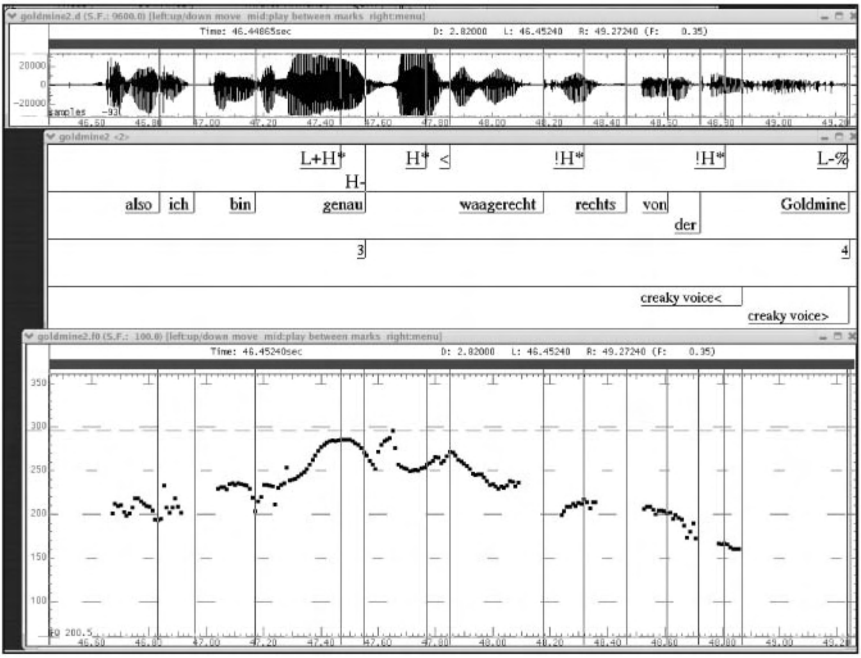


FIGURE 3.1 Fo contour and label tiers of the utterance 'Also ich bin genau waagerecht rechts von der Goldmine' (*Well, I'm exactly in a horizontal line to the right of the gold mine*) (adapted from Grice and Benzmüller 1995).

of each tone label to the positions of the word labels. Thus, a pitch accent label falling within the bounds of an annotated word is taken to highlight that word. The stressed syllable of the word must be identified separately, since the convention is not to mark stress in the orthographical representation.

We assume that most databases will be syllabified or even be annotated at the segmental level, in which case the syllable bearing a given accent can be found by relating the time stamp of the tone label to that of the syllables or segments. However, should a database have no such annotations we suggest explicitly marking cases where word stress is optional, as in the word for 'coffee' which may be pronounced *'Kaffee* or *Kaf'fee* (Duden 2000). As above, the IPA stress mark (') is inserted in the orthographical string before the stressed syllable. We also suggest marking cases where the pitch accent occurs on a syllable other than the primary lexical stress. In (3*a*) below, there are no stress marks, since the accent is on the lexically stressed syllable *zwan*. This is not the case in (3*b*), where the accent is shifted backwards onto *hun*.[3] The IPA stress mark is given in (3*b*) as it would appear in the label file.

(3) (*a*) hundertzwanzig
hundred and twenty
(*b*) 'hundertzwanzig Mann
hundred and twenty men

(adapted from Giegerich 1985: 218)

Since the information as to which syllable is accented is gained only indirectly, it is important that the label for a pitch accent is placed within the bounds of the associated syllable. This is not a problem if the pitch accent peak or valley occurs within the syllable. In this case, the label is placed on the peak or valley. However, if the peak or valley occurs outside the syllable, we follow the MAE_ToBI conventions: the label has to be placed within the associated syllable and the actual peak or valley is marked with a '>' or '<' label, depending on whether it occurs before or after the associated syllable, respectively. In German the peak of a L+H* pitch accent is generally reached late in an accented syllable, often even later (especially if the accented syllable is short). This is also occasionally true for simple H* pitch accents. To avoid a situation where labellers have to decide on the location of syllable boundaries, the '<' label should be used in cases where the peak is somewhere in the

[3] In their study of American English, Shattuck-Hufnagel *et al.* (1994) found that the perceived shift of prominence from the lexical stress to an earlier syllable in rhythmic clash contexts is due to substantial Fo movement rather than any durational increase on the prominent syllable. It is thus more appropriate to call this type of shift a 'pitch accent shift' rather than the traditional term 'stress shift'.

vicinity of the syllable boundary (the same consideration holds for L* and valleys). An example of the '<' symbol can be found in Figure 3.1.

Transcriber confidence as to the accuracy of individual labels is captured by a '?' flag after uncertain labels, and a '$' flag where the example is perceived to be a prototypical realization of a given category.

3.3.2. *Pitch accents*

The six basic pitch accents in GToBI are described below. The H and L tones are described as high or low relative to a speaker's pitch range which can be thought of as having a topline as an upper limit and a baseline as a lower limit. As a rule of thumb, tones which are perceived as high are roughly in the top three-quarters of the range, those perceived as low in the bottom quarter.

- H* 'peak accent'

A canonical H* syllable is perceived as relatively high and may be preceded by a shallow rise.

- L+H* 'rise from low up to peak accent'

Here, as in H*, the accented syllable is perceived as high. It is preceded by a syllable with a low pitch target which leads to a sharp rise in (or a jump up to) the accented syllable. The peak is often late in the accented syllable (cf. Adriaens 1991; Grabe 1998).

- L* 'low accent'

The L* syllable is a local pitch minimum low in the speaker's range. It may be preceded by a shallow fall.

- L*+H 'valley accent plus rise'

Here a low target within the accented syllable is followed by a rise, starting late in the accented syllable and reaching its peak on the next syllable (or sometimes later). In contrast to L+H*, the perceived pitch of the accented syllable is low.

- H+L* 'step-down from high to low accent'

The accented syllable is low with a valley clearly at or very near the bottom of the speaker's range. It is preceded by a high pitch target which generally occurs on the syllable immediately preceding the accented syllable.

- H+!H* 'step-down from high to mid accent'

As in H+L*, the accented syllable is preceded by a higher pitch on the immediately preceding syllable. However, the accented syllable is not low, but rather around the middle of the range between the 'H+' peak and the speaker's baseline. If H+!H* is immediately followed by a low boundary tone there is a continuous fall from the preaccented syllable, through the accented syllable up to the next stressed syllable, if there is one, otherwise to the final syllable of the phrase.

The absence from the GToBI inventory of a H*+L accent will be discussed in detail in Section 3.4.2.

The six basic pitch accents may additionally be scaled within a modified pitch range. The most common modification involves a lowering of the topline by a process of downstep, shifting the pitch of H tones downwards. When this occurs, the affected H tone is marked by a preceding '!' symbol. This is what happens in the H+!H* pitch accent. However, any H tone from the basic set of pitch accents can undergo downstep (e.g. !H*, L*+!H, and so on). It is important to note that within an intonation phrase all H tones following a downstep are scaled within the same reduced range. These following H tones are not especially marked for range, unless there is a subsequent step down. When there is more than one step down, i.e. when downstep occurs in sequence, then it is transcribed separately at each step, such as in the second phrase of Figure 3.1, ... *WAAgerecht RECHTS von der GOLDmine* ('... in a horizontal line to the right of the gold mine'): 'H* !H* !H* L-%'. In the more phonologically oriented autosegmental-metrical models, downstep is either triggered automatically by a particular pitch accent type, as in Pierrehumbert's original English model (1980), or treated as an optional operation (Gussenhoven 1984). Since GToBI is more surface-oriented, it is simply flagged explicitly each time there is a step down in pitch.

When a new phrase is begun after a phrase containing a downstep, the pitch range is reset. In certain cases, a sequence of downsteps may be followed by a reset within a phrase, usually just before the nuclear accent. That is, after a sequence of steps down, there is a step up to the peak on the nuclear syllable. This happens in English (Ladd 1983: 735, example 4(b)) and has also been attested in Southern German (Truckenbrodt 1998, 2000). GToBI makes use of an upstep '^' symbol to capture such cases, as exemplified in Figure 3.3 (Section 3.4.1 (i)). Phrases may also contain sequences of pitch accents where each accent involves a step up in pitch (e.g. in emphatic speech). Such contours are transcribed by Selting (1995) as globally rising. In GToBI each accent of such a sequence is marked with a '^' symbol.

The explicit marking of upstep, which distinguishes GToBI from other autosegmental-metrical accounts, is not only used to indicate a step up within a sequence of pitch accents, it also describes a step up to a boundary tone, as will be shown in Section 3.3.3 below.

3.3.3. *Boundary tones*

GToBI differentiates two levels of phrasing: the (minor) intermediate phrase and the (major) intonation phrase. Each intonation phrase contains at least one intermediate phrase, and each intermediate phrase contains at least one pitch accent.[4] The edge tones for these phrases determine the contour from the last tone of the last pitch accent until the end of the phrase. There are three intermediate phrase edge tones:

- L-

L- constitutes an Fo minimum low in the range.

- H-

H- has roughly the same Fo value as the peak corresponding to the most recent H tone in the phrase. Given enough distance, there is a plateau between a nuclear H tone and the end of the phrase.

- !H-

A H- tone can also be downstepped in relation to a previous accentual H tone. This most commonly occurs in calling contours.

Theoretically, there is the possibility of a fourth intermediate phrase boundary type, ^H-, although we have not to date found examples which would unambiguously distinguish it from H- like others.

The target for the intermediate phrase edge tone is often reached at a postnuclear stressed syllable (if there is one) and extends up to the beginning of the last syllable of the phrase. The tendency for the intermediate phrase boundary tone to align with postnuclear stressed syllables is reported on in Grice and Benzmüller (1998) and is evidence for it being a phrase accent, as discussed in Grice *et al.* (2000). Phrase accents are tones which function as edge tones but can also associate with stressed syllables or other tone-bearing units. In GToBI there is an option to explicitly transcribe this association with a separate L(*) or H(*) label, whereby the star in brackets denotes the secondary nature of the postnuclear prominence (see Figure 3.4 in Section 3.4.1 (i) for an example of usage).

[4] Exceptions to this rule are the so-called 'intonational tags', which can be regarded as enclitic tone units without pitch accents.

An intonation phrase (IP) edge never occurs without a preceding intermediate phrase (ip) edge. Their tones are therefore listed below as combinations. The new GToBI presented here has simplified the boundary tone notation in order to make it phonetically more transparent. For example, in cases where the IP and ip boundary tones would represent the same pitch level, only one tone is transcribed: H-% in the new GToBI instead of the original H-L% (see below for explanation), and L-% instead of L-L%. The description of the canonical shapes given below assumes a distance of at least two syllables from the final pitch accent to the end of the phrase.[5]

- H-% 'plateau'

The main difference between H- and H-% is not tonal, but rather relates to perceived boundary strength, as encoded by the labels 3 and 4 in the parallel break index tier (see below). The similarity between the two contours is captured by the use of only one H tone.

There have been several different ways of describing such a plateau at phrase boundaries. Grabe (1998), for example, suggests the transcription 0% for a contour that more or less stays the same from the end of the last pitch accent to the boundary. The problem with this transcription is that the unmarked boundary tone does not directly encode whether the phrase ends low or high. Its value depends on which accent precedes it. The original GToBI transcription of a plateau was H-L% (with automatic upstep on the L% tone). Since using an L tone to represent mid or high pitch was considered counter-intuitive and difficult to learn, the new GToBI transcription eliminates the L tone altogether. The combined label H-% has the advantage of directly encoding the phrase final pitch height without syntagmatic reference to preceding pitch accents. This makes the system easier to learn and more straightforward for database access.

- H-^H% 'plateau followed by sharp rise at the end of the phrase'

The upstepped H% component causes a sharp rise in the last syllable of the phrase, often to a point very high in the speaker's range.

- L-H% 'low followed by rise to mid at end of phrase'

This edge tone combination accounts for a final fall-rise contour if it is preceded by a H tone, and otherwise simply for a low stretch with a rise to mid on the last syllable.

- L-% 'low stretch which may be followed by drop to extra low'

[5] If the final accented syllable is closer to the boundary, then much of the shape is lost owing to lack of time for the realization of the individual tones.

The main difference between L- and L-% is perceived boundary strength. In addition, L-% is generally lower than L-. A final drop at the end of the phrase is possibly due to factors such as final lowering which is little understood in German and is not necessarily confined to the final syllable. We do not distinguish between several degrees of lowness at the IP boundary.[6]

- %H 'initial high boundary'

GToBI also provides the option of marking a high intonation phrase onset with a %H initial boundary tone. A mid or low tone is not explicitly marked.

The following boundary tone combinations are not in GToBI at present:

- L-^H%

In principle, this combination could be used to transcribe a low stretch with a rise to extra high on the final syllable. However, we do not have clear examples of this contour as distinct from H-^H%.

- H-L%

This combination was used in the original GToBI to describe a level contour (with automatic upstep of the L% after a H- phrase accent). Since upstep is now marked explicitly, H-L% could be used to describe a fall to low after a high plateau. Although this contour is not attested in Standard German, it has been reported in an East Phalian dialect (Kerckhove 1948: 63) and in dialects of the Palatinate (Peters 2001*a,b*).

The other three logical possibilities for boundary tone combinations are captured by simpler ones, already given above in the inventory: H-H% and H-^L% can be equated with H-%, and L-^L% would describe a contour very much resembling L-H%. Further nuances in the description of tonal movements at intonation phrase boundaries have not yet proved necessary in the sense that GToBI already captures all the IP phrase final contours reported to have distinct meanings or functions.

3.3.4. *Break indices*

The break index tier is based on MAE_ToBI, where in the default case '3' and '4' coincide with intermediate phrase and intonation phrase boundaries

[6] This means that we do not distinguish between L-% and L-L% to transcribe the presence or absence of a final drop. For the moment, we leave open whether this distinction is functionally motivated for a transcription system of Standard German (cf. Peters (2001*a*) where the two transcriptions are used to express a difference in the tonal alignment of postnuclear falls in a regional variety of German).

respectively. GToBI does not explicitly mark the break index unless it deviates from this. In such cases, there are three options: the label '4-' is used for cases where a phrase boundary is perceived, but where it is unclear as to the level of phrasing. GToBI distinguishes between two mismatches in tonal and rhythmic structure which are both encoded in MAE_ToBI with index 2: a rhythmic break with tonal continuity, '2r', e.g. a rhetorical or hesitation pause; and a tonal break with rhythmic continuity, '2t', e.g. a perceived boundary without a pause but with a tonal contour not attributable to the accents in the phrase. This often occurs in fast speech. GToBI does not include a simple break index 2 in its inventory. Break indices below level 2 are not dealt with.

A summary of the proposed annotations can be found in the Appendix.

3.3.5. *Commonly occurring nuclear contours*

Schematic representations and textual examples of commonly occurring nuclear contours are given in Table 3.1, along with a suggested context in which such an utterance might be produced. In the schematic contours, extra heavy lines represent accented syllables, heavy lines postnuclear stressed syllables, and dotted lines the baseline of the speaker's pitch range. The line drawings provide a maximally long contour, assuming that the nucleus is followed by at least one postnuclear stress (the heavy line) and at least one other syllable after that. Most of the contour types have at least one example with a postnuclear stress. However, since the examples were chosen because they are representative in terms of their pragmatic interpretation and not because of their rhythmic structure, some do not correspond to the maximal contour. For instance, the rise (3*a*) has two example sentences: *Tauschen Sie auch **BRIEF**MARken?* ('Do you also exchange stamps?') and *Von wem ich das **HA**be?* ('From whom I have it?'). In the first, the extra heavy line corresponds to the nucleus, ***BRIEF***, the heavy line to the postnuclear stress *MAR*. In the second, the nuclear syllable ***HA*** is followed by only one syllable. Since there are not enough segments for the realization of the rise-plateau-rise shape, the pitch simply rises directly from ***HA*** to the end of the phrase. In cases where the nuclear syllable is final in the phrase, especially if the coda contains voiceless obstruents, the contour may be truncated. This is particularly true in the case of falling contours (see Grabe 1998).

The contexts provided in the table contain pragmatic interpretations referring to specific examples; they should not be taken as abstract meanings for given contours. If syntactic information is given, then it is simply that the pattern may be regarded as neutral for a particular syntactic construction. It does not imply any more than this. We do not distinguish between linguistic and paralinguistic

Table 3.1 Commonly occurring German nuclear contours and examples of their usage

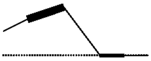
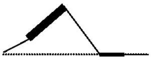
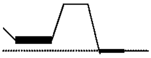
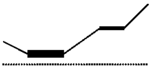
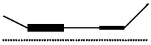
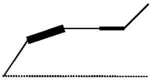
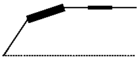
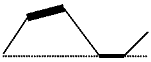
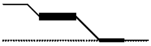
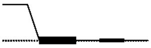
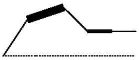
		GToBI	Schematic contour	Context	Example
Fall	1a	H* L-%		Neutral statement	Mein **ZAHN** tut WEH.[1] *My tooth hurts*
				Neutral W-question	Wo hast du den **WA**gen gePARKT?[1] *Where did you park the car?*
	1b	L+H* L-%		Contrastive assertion	Schon der Ver**SUCH** ist STRAFbar![2] *Even to attempt is an offence!*
Rise-fall (Late peak)	2	L*+H L-%		Self-evident assertion	Das **WEISS** ich SCHON![6] *I already know that!*
				Emotionally committed or sarcastic assertion	Der Blick ist ja **FA**belhaft![3] *The view is fantastic!*
Rise	3a	L* H-^H%		Neutral yes/no-question	Tauschen Sie auch **BRIEF**MARken?[1] *Do you also exchange stamps?*
				Echo question	Von wem ich das **HA**be?[2] *From whom I have it?*
	3b	L* L-H%		Indignation	**DOCH**! *It is!*
				Answering phone	**BEC**kenBAUer?[4]
	3c	(L+)H* H-^H%		Follow-up question	. . . oder ist Ihr **BRU**der HIER?[5] *. . . or is your brother in?*
Level	4	(L+)H* H-(%)		Incompleteness	**AN**dererSEITS . . .[6] *But then again . . .*

TABLE 3.1 (*Continued*)

		GToBI	Schematic contour	Context	Example
				Ritual expression	Guten **MOR**gen![3] *Good morning!*
Fall-rise	5	(L+)H* L-H%		Polite offer	Mögen Sie **ROG**genBRÖTchen?[1] *Would you like rye rolls?*
Early peak	6a	H+!H* L-%		Established fact	Hab' ich mir schon ge**DACHT**[7] *That's what I thought*
	6b	H+L* L-%		Soothing/ polite request	Nun er **ZÄH**le doch MAL![2] *Just tell me about it!*
Stylized step down	7	(L+)H* !H-%		Calling contour	**BEC**kenBAUer!

Note: Examples are taken from [1] Féry (1993), [2] von Essen (1964), [3] Fox (1984), [4] Ladd (1996, adapted), [5] Moulton (1962), [6] Pheby (1984), and [7] Grice and Benzmüller (1995). Capitals in bold face indicate nuclear syllables, plain capitals postnuclear stresses.

functions of intonation, since it has been shown that both types of function can be expressed by discrete means such as the choice of pitch accent and boundary tones (Scherer *et al.* 1984). We therefore include information as to speaker attitude or affect, where this helps to clarify the context in which an utterance might be spoken. We discuss each section of the table separately below.

Fall: In the autosegmental literature there is only one type of fall. In GToBI there is a simple fall, represented as H* L-%, and a fall preceded by a sharp rise. The latter is represented with a leading L tone, thus L+H* L-%. Although this combination does not necessarily signal contrast, it may do so (especially with a wide pitch range), as in the example given in Figure 3.2.

Rise-fall: The rise-fall, represented as L*+H L-%, differs in timing from L+H* L-%; in the former, the rise begins later in the accented syllable than in the latter. In the former the accented syllable sounds low whilst in the latter it is clearly high.

Rise: In the early autosegmental-metrical literature there are at most two different types of rise. In GToBI there are two starting on a low pitch, L* L-H% and L* H-^H%, where the endpoint of the second is higher than the

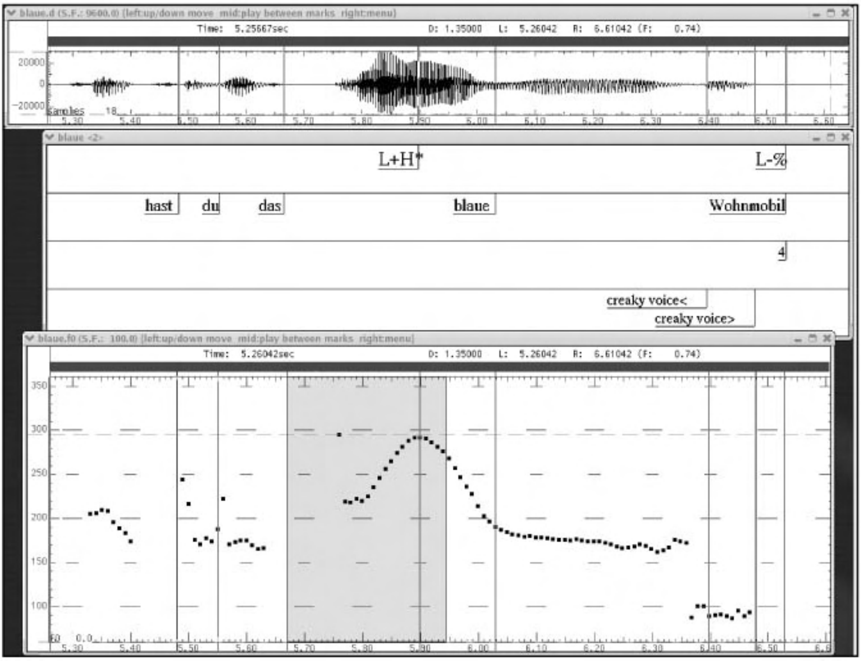

blaue.d (S.F.: 9600.0) (left:up/down move mid:play between marks right:menu)
Time: 5.25667sec
D: 1.35000 L: 5.26042 R: 6.61042 (F: 0.74)
20000
0
−20000
Samples 18
5.30
5.40
5.50
5.60
5.70
5.80
5.90
6.00
6.10
6.20
6.30
6.40
6.50
6.60
blaue <2>
L+H*
L-%
hast
du
das
blaue
Wohnmobil
4
creaky voice<
creaky voice>
blaue.f0 (S.F.: 100.0) [left:up/down move mid:play between marks right:menu]
Time: 5.26042sec
D: 1.35000 L: 5.26042 R: 6.61042 (F: 0.74)
350
300
250
200
150
100
50 0.0

FIGURE 3.2 F0 contour of L+H* L-% on 'Hast du das BLAUe WOHNmobil?' (*Do you have the blue caravan?*) (adapted from Grice and Benzmüller 1995); for clarification purposes the shaded area marks the nuclear syllable *BLAU*.

first. There is also a rise where the accented syllable is mid, with or without a steeply rising onglide, i.e. (L+)H* H-^H%.

Level: Contours ending in a level or sustained pitch are barely mentioned in the literature. According to Féry, a L*+H rise can be followed by a level stretch which is not given any explicit transcription; it is assumed that the pitch of the trailing tone is continued until the end of the phrase (i.e. progredient intonation). She therefore does not distinguish between rises and level nuclear patterns, claiming that 'As a matter of fact, rising tones and progredient intonation cannot be kept apart' (1993: 89). GToBI marks a level contour with or without a steeply rising onglide as L+H* H-% or H* H-% respectively.

Fall-rise: Generally GToBI represents fall-rises as (L+)H* L-H%. In principle, it is also possible to mark a 'high fall-rise' (where the pitch between the two peaks does not drop to low) as (L+)H* !H-^H%, although it is unclear whether this distinction is really necessary.

Early peak: GToBI has two early peak contours: H+!H* and H+L*. The former is the early peak contour referred to by Kohler with the meaning 'established fact', and also the one transcribed by Féry as H+H*+L. Von Essen claims that this pattern can signal finality even on unfinished parts of an utterance, and points out that it is often used in radio announcements. This matches Féry's claim that this pattern is frequently used by TV reporters. The schematic contour for H+!H* allows for a postnuclear stressed syllable which is not present in the example (selected from a spontaneously produced corpus). An example such as *Sie hätte ja* ***LÜ****gen KÖNnen* ('She even could have lied') would have the syllable *kön* on the second heavy line. The other early peak contour, H+L*, is what von Essen describes as signalling a fatalistic tone. It can also be used for soothing or polite requests, as in the example given in Table 3.1. Early peak contours will be discussed in more detail in Section 3.4.1(i).

Stylized step down: The stylized step down (or calling contour) is represented as (L+)H* !H-%. Here the phrase accent !H- occurs on a stressed syllable if there is one. The prominent syllable upon which the step down occurs is optionally marked with '!H(*)'. The multiple uses of this contour in German have been described at length by Gibbon (1976, 1998).

3.4. GToBI COMPARED WITH OTHER AM ACCOUNTS

Table 3.2 provides correspondences between GToBI and the models of Wunderlich (1988), Uhmann (1991), and Féry (1993). The gaps in the table indicate that GToBI makes more distinctions than the other models. The

TABLE 3.2 German nuclear contours: three models compared with GToBI

		Wunderlich	Uhmann	Féry	GToBI
Fall	1 a	H* L	H*+L L%	H*+L	H* L-%
	1 b				L+H* L-%
Rise-fall (Late peak)	2		L*+H L%	L*+H+L	L*+H L-%
Rise	3 a	L* H H%	L*+H H%	L*+H	L*(+H) H-^H%
	3 b	L* H%			L* L-H%
	3 c				(L+)H* H-$^{\wedge}$H%
Level	4			L*+H	(L+)H* H-(%)
Fall-rise	5		H*+L H%	H*+L H%	(L+)H* L-H%
Early peak	6 a			H+H*+L	H+!H* L-%
	6 b	%H L* L			H+L* L-%
Stylized step down	7			H*+M	(L+)H* !H-%

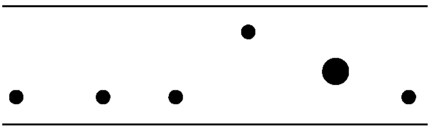
increased expressivity of GToBI is due to a number of different factors. It has leading tones, thus enabling distinctions to be made between, for example, a plain fall and a fall with a preceding onglide. It allows for a phrase accent as well as an intonation phrase boundary tone, whereas the other models have only one edge tone. Each of these issues will be dealt with separately below.

3.4.1. *Leading tones*

In GToBI it is possible to have either H or L as leading tones. This means that the pitch before an accented syllable may be transcribed as high, in which case the contour is referred to as an early peak, or low, in which case there is a rise up to the accented syllable, referred to as a rising onglide.

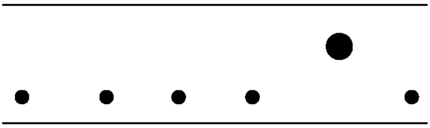
(i) Early peak contours: We have seen ample evidence for the existence of early peak contours, translatable into contours with a H leading tone. Kohler's early peak, exemplified in (4*a*), provides us with such a case. It is contrasted with (4*b*), which he refers to as a medial peak.

(4) (*a*)



Sie hat ja ge- LO- gen
She had actually LIED
'She actually LIED'

(*b*)



Sie hat ja ge- LO- gen
She had actually LIED
'She actually LIED'

(adapted from Kohler 1995: 123)

Moreover, Kohler performed perception tests which clearly indicated that high pitch on the preaccentual syllable (i.e. the one immediately before the accented syllable) is distinctive. That is, the early peak contour, which signals that information is old, is linguistically distinct from a medial peak signalling new information. GToBI captures this distinction by representing the early peak contour with a leading H tone, implying that the peak is *before* the accented syllable, and the medial peak contour with a H* (possibly with a leading L tone, see 3.4.1 (*ii*)), implying that the peak is *on* the accented syllable.

We have also seen that within the levels approach of Isačenko and Schädlich an early peak is represented as a 'preictic' fall, schematized in (5*a*) below, in contrast to a 'postictic' fall (5*b*).

(5*a*) preictic (5*b*) postictic
die ↓KINder die KIN ↓ der
the children the children
(Isačenko and Schädlich 1966: 60)

However, GToBI does not only have one early peak contour, but rather two: H+!H* and H+L*. Grabe (1998: 89f.) argues that this distinction can be interpreted as one between total and partial downstep of her basic H*+L pitch accent (although she claims the distinction is gradual). As we have seen in 3.3.5, von Essen (1964) describes both types of early peak contour, attributing different usages to each of them. Figures 3.3 and 3.4 illustrate spontaneous utterances of H+!H* L-% and H+L* L-% contours.

(ii) Rising onglides: Selting (1995) claims that a distinctive local pitch pattern begins on an accented syllable and is extended over the following unaccented syllables up to, but not including, the next accent. Thus the domain of the accent is related to the domain of the Abercrombian foot, or the 'Takt' (Pheby 1975;

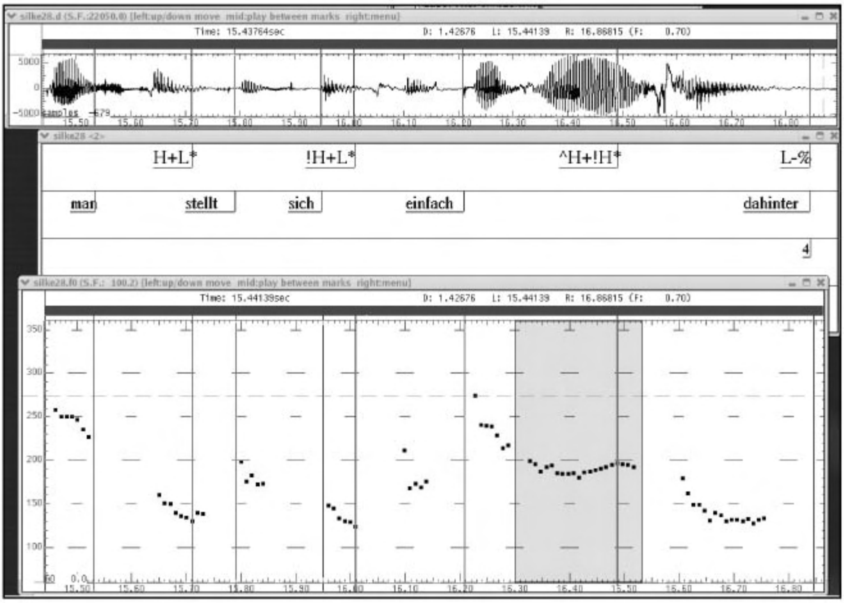


FIGURE 3.3 Fo contour of H+!H* L-% on 'Man stellt sich einfach daHINter' (*You just queue up behind it*); the shaded area marks the nuclear syllable *HIN*.

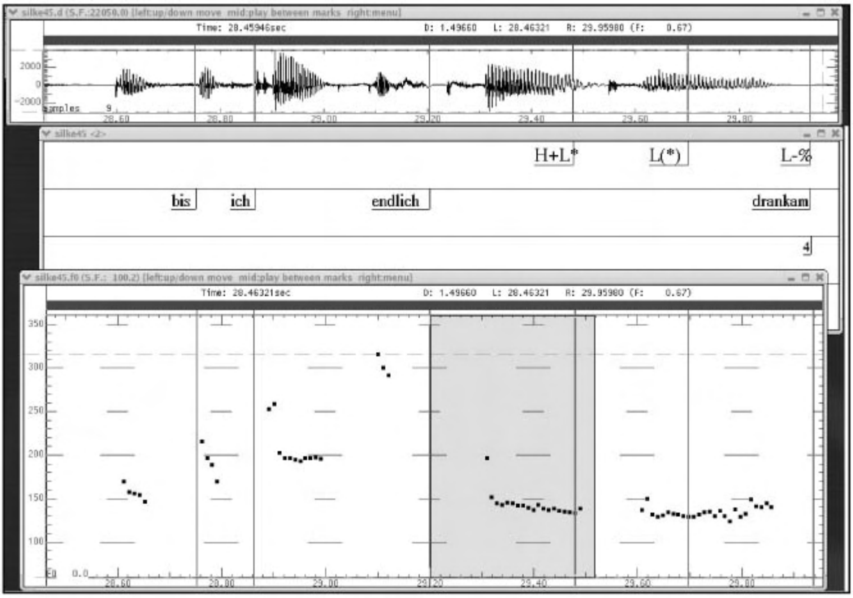

silke45.d (S.F.:22050.0) (left:up/down move mid:play between marks right:menu)
Time: 28.45946sec
D: 1.49660 L: 28.46321 R: 29.95980 (F: 0.67)
2000
0
-2000
samples 9
28.60
28.80
29.00
29.20
29.40
29.60
29.80
silke45 <2>
H+L*
L(*)
L-%
bis
ich
endlich
drankam
4
silke45.f0 (S.F.: 100.2) (left:up/down move mid:play between marks right:menu)
Time: 28.46321sec
D: 1.49660 L: 28.46321 R: 29.95980 (F: 0.67)
350
300
250
200
150
100
F0 0.0

FIGURE 3.4 Fo contour of H+L* L-% on '. . . bis ich endlich DRANKAM' (. . . *until it finally was my turn*); the shaded area marks the nuclear syllable *DRAN*.

Kohler 1977), as it is referred to in German, and identical to Gussenhoven's (1990) 'Tonal Association Domain'. The domain of Selting's local pitch pattern is thus equivalent (at least for the final accent in a phrase) to the domain of the nuclear tone in the British School. This means that the pitch immediately prior to an accented syllable is excluded from the analysis of that accent. However, in the early configurations-based approaches there are descriptions of patterns in which the pitch of the immediately prenuclear syllable is important for the interpretation of the contour. In example (6) from Fox (1984: 19), where the nuclear accent is on *KOMMST*, the British School is forced to analyse the nuclear tone as level. This would also be the case in Selting's approach.

(6) ● ● ● ●
Wenn du morgen KOMMST . . . (fahren wir zusammen)
If you tomorrow COME . . . (go we together)
'If you COME tomorrow . . . (we can go together)'

However, the movement from *gen* to *KOMMST* is clearly perceived as rising or as a jump up to the nuclear syllable. The specification of a level pitch on *KOMMST* is not enough to capture the essence of this contour. Fox states this clearly:

> [. . .] an important characteristic of this pattern is the *jump up* to the high level pitch of the nucleus. The nucleus must always be at a higher pitch than the immediately

preceding syllable. If the preceding head is high, its pitch must fall towards the end to allow for the jump up, hence the lower pitch given to *morgen* [. . .]

(1984: 19f., italics as in original)

This jump up is also represented in the early levels-based approach of Isačenko and Schädlich as a preictic rise, as schematized in (7).

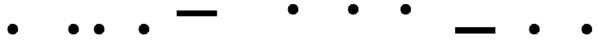
(7) die ↑ KINder
the children (Isačenko and Schädlich 1966: 60)

GToBI captures this tonal movement by the leading L tone in a L+H* pitch accent.

3.4.2. *Levels of phrasing and phrase accents*

Autosegmental-metrical accounts of German, as in much of the traditional literature, generally restrict the levels of phrasing to only one—the intonation phrase (e.g. Uhmann 1991; Grabe 1998). The AM exception is Féry (1993), who, like Pierrehumbert for English and GToBI for German, assumes intermediate phrases as well. Among the auditory approaches, von Essen's is the one which could be interpreted as allowing for an additional smaller level of phrasing. His unit of analysis is the rhetorical phrase. He distinguishes between two types of rhetorical phrase: a major one with a nucleus (or 'Schwerpunkt') and a minor one without a nucleus. When an utterance contains more than one phrase, he claims it is the last one which contains a nucleus. This can be seen in example (8) below:

(8)



Ich habe ge- TAN | was mir be- **FOH-** len war.
I have done what me ordered was
'I DID what I was ORDERED to do.' (von Essen 1964: 38)

In a GToBI analysis, example (8) would be divided into two intermediate phrases (ip) forming one intonation phrase (IP), as in (9) below:

(9) [[Ich habe geTAN]ip [was mir beFOHlen war]ip]IP

However, although the analyses of von Essen and GToBI appear to be similar, there is one important difference: von Essen distinguishes the two types of phrase according to their pitch contours (progredient vs. terminal and interrogative). GToBI, by contrast, provides two different, hierarchically structured domains of phrasing, which are independent of specific pitch contours.

The only restriction on contours at a given boundary is the number of tones available to capture them. At an intermediate phrase boundary the phrase accent is the only edge tone available. At an intonation phrase boundary there are always two edge tones: the phrase accent and the intonation phrase boundary tone. In the latter case there is one more tone available, which allows for a more complex tonal contour to be captured.

Féry dispenses with the intermediate phrase boundary tone, which she also refers to as the phrase accent. She claims that the phrase accent has two functions: to control the pitch movement between the nuclear accented syllable and the boundary tone of an intonation phrase, and to mark the boundary of an intermediate phrase (1993: 79), and argues that the trailing tone of the phrase-final pitch accent takes over both of these functions. However, such an approach would make it difficult, if not impossible, to disambiguate between a sequence of prenuclear bitonal pitch accents (which Féry (1993: 120) explicitly allows for) and a sequence of intermediate phrases, each with one bitonal pitch accent (which are also possible in her model).

The transcription of nuclear falls as H*+L, as in Féry and Uhmann, or as (L+)H* L- as in GToBI is still a controversial issue. Grice and Benzmüller (1998) found that the 'elbow' (the point at which the pitch reaches the baseline) after a medial peak accent differed according to the number of unstressed syllables between the nuclear syllable and the next postnuclear stressed syllable; the further away the stressed syllable, the later the baseline was reached. In fact, in 94 per cent of fall-rises and in 91 per cent of falls, the baseline was reached precisely on the postnuclear stressed syllable. This is taken as strong evidence for the analysis of those patterns as H* L-H% and H* L-% respectively, as opposed to H*+L H% and H*+L L%.

However, there are possibly contours which have not yet been investigated where the elbow is aligned differently, for example at a relatively constant distance from the H* peak, indicating that the L tone is part of a bitonal H*+L pitch accent (this could be the case in structures displaying narrow focus or contrast, as several examples in Uhmann (1991) suggest). We assume that there are dialectal differences in the position of the elbow. This is a common phenomenon in other languages, as is the case in Greek, Romanian, and Hungarian where the question tune has the same tonal structure but has a different association of the phrase accent tone, depending on the dialect. As a matter of fact, in an analysis of Bern Swiss German, Fitzpatrick-Cole (1999) has found evidence for the association of an edge tone (analysed as an intonation phrase boundary tone) to a lexically stressed syllable. In that dialect it appears that the tone associates with the final stressed syllable in the phrase rather than the immediately postnuclear one.

3.5. PROSPECT

GToBI is an annotation scheme, which should be regarded as a tool for research into the phonological structure of German intonation. Although its descriptive inventory is certainly richer than any other AM model of German intonation, the distinctions GToBI offers are all motivated by independent studies, either auditory or instrumental. This is hardly surprising since the aim of an annotation scheme is to be as flexible as possible in order to capture all relevant empirically observed patterns, even if it turns out later that part of the description is redundant, in that some of these patterns are derivative or make reference to gradient features rather than categorical ones.

APPENDIX: SUMMARY OF GToBI LABELS

pitch accents	
H*	peak accent
L+H*	rise from low up to peak accent
L*	low accent
L*+H	valley accent plus rise
H+L*	step-down from high to low accent
H+!H*	step-down from high to mid accent
phrase accents	
H(*)	postnuclear prominence on high syllable (optional)
L(*)	postnuclear prominence on low syllable (optional)
boundary tones	
H-	high intermediate phrase boundary
L-	low intermediate phrase boundary
!H-	downstepped intermediate phrase boundary
H-%	high plateau at intonation phrase boundary
H-^H%	high plateau followed by sharp rise at end of phrase
L-H%	low followed by rise to mid at end of phrase
L-%	low stretch which may be followed by drop to extra low
%H	initial high boundary
scaling of pitch range	
!	downstep of H pitch accents and H intermediate phrase boundary tones
^	upstep of H pitch accents and H boundary tones
break indices	
2r	rhythmic break with tonal continuity
2t	tonal break with rhythmic continuity

3	intermediate phrase boundary
4	intonation phrase boundary
4-	uncertain level of phrasing
other labels	
'	word stress
>	peak or valley before syllable associated with an accent
<	peak or valley after syllable associated with an accent
?	uncertain label
$	prototypical example

REFERENCES

Adriaens, L. M. H. (1991), 'Ein Modell deutscher Intonation: eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text', Ph.D. dissertation (Eindhoven University of Technology).

Altmann, H., Batliner, A., and Oppenrieder, W. (1989) (eds.), *Zur Intonation von Modus und Fokus im Deutschen* (Tübingen: Niemeyer).

Arvaniti, A., Ladd, D. R., and Mennen, I. (1999), 'What is a starred tone?', in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon* (Cambridge: Cambridge University Press), 119–31.

Beckman, M. E., and Ayers-Elam, G. (1997), 'Guide to ToBI Labelling', Electronic text and accompanying audio example files available at http: //ling.ohio-state.edu/Phonetics/E_ToBI/etobi_homepage.html.

—— and Hirschberg, J. (1994), *The ToBI Annotation Conventions.* Online MS. Available at http://www.ling.ohio-state.edu/tobi/ame_tobi/annotation_conventions.html.

——, —— and Shattuck-Hufnagel, S. (this volume Ch. 2), 'The original ToBI system and the evolution of the ToBI framework'.

Benzmüller, R. and Grice, M. (1997), 'Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI', *Phonus* (Saarbrücken University), 3: 9–34.

——, —— (1998), 'The nuclear accentual fall in the intonation of Standard German', in *ZAS Papers in Linguistics: Papers on the Conference 'The word as a phonetic unit'*, (Berlin), 79–89.

Crystal, D. (1969), *Prosodic Systems and Intonation in English* (Cambridge: Cambridge University Press).

Duden (2000), *Aussprachewörterbuch* (Duden Volume 6), (4th edn., Mannheim: Dudenverlag).

Essen, O. von (1964) (2nd edn.), *Grundzüge der hochdeutschen Satzintonation* (Ratingen: Henn, 1st edn. 1956).

Féry, C. (1993), *German Intonational Patterns* (Tübingen: Niemeyer).

Fitzpatrick-Cole, J. (1999), 'The Alpine Intonation of Bern Swiss German', in *Proceedings of the ICPhS99* (San Francisco), 941–4.

Fox, A. (1984), *German Intonation: An Outline* (Oxford: Clarendon Press).

Gibbon, D. (1976), *Perspectives of Intonation Analysis* (Bern: Lang).

—— (1998), 'Intonation in German', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems. A survey of 20 languages* (Cambridge: Cambridge University Press), 78–95.

Giegerich, H. J. (1985), *Metrical Phonology and Phonological Structure. German and English* (Cambridge: Cambridge University Press).

Grabe, E. (1998), *Comparative Intonational Phonology: English and German* (MPI Series in Psycholinguistics 7) (Wageningen: Ponsen and Looijen).

Grice, M. and Benzmüller, R. (1995), 'Transcription of German Intonation using ToBI-Tones—The Saarbrücken System', *Phonus* (Saarbrücken University), 1: 33–51.

——, —— (1998), 'Tonal Affiliation in German Falls and Fall-rises', poster presented at the 6th Conference on Laboratory Phonology, York, UK, 2–4 July.

——, Ladd, D. R., and Arvaniti, A. (2000), 'On the Place of Phrase Accents in Intonational Phonology', *Phonology*, 17/2: 143–85.

——, Reyelt, M., Benzmüller, R., Mayer, J., and Batliner, A. (1996), 'Consistency in Transcription and Labelling of German Intonation with GToBI', in *Proceedings of the Fourth International Conference on Spoken Language Processing* (Philadelphia), 1716–19.

Gussenhoven, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

—— (1990), 'Tonal Association Domains and the Prosodic Hierarchy in English', in S. Ramsaran (ed.), *Studies in the Pronunciation of English* (London: Routledge).

Halliday, M. A. K. (1967), *Intonation and Grammar in British English* (The Hague: Mouton).

Isačenko, A. V. and Schädlich, H. J. (1966), 'Untersuchungen über die deutsche Satzintonation', *Studia Grammatica*, 7: 7–64.

Kerckhove, M. van de (1948), 'Intonationssystem einer Mundart', *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft*, 52–65.

Kohler, K. J. (1977), *Einführung in die Phonetik des Deutschen* (Grundlagen der Germanistik 20), (Berlin: Schmidt, 2nd edn. 1995).

—— (1991), 'Terminal Intonation Patterns in Single-accent Utterances of German: Phonetics, Phonology, and Semantics', *AIPUK* (Kiel University), 25: 115–85.

Ladd, D. R. (1983), 'Phonological Features of Intonational Peaks', *Language*, 59: 721–59.

—— (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

Moulton, W. G. (1962), *The Sounds of English and German* (Chicago: University of Chicago Press).

Peters, J. (2001*a*), 'Postnukleare Tonhöhengipfel in der Vorderpfalz und in Mannheim', Ms. (University of Potsdam).

—— (2001*b*), 'Frageintonation in der Pfalz. Eine Reanalyse des Güntherodt-Korpus' MS University of Potsdam.

Pheby, J. (1975), *Intonation und Grammatik im Deutschen* (Berlin: Akademie-Verlag).

PHEBY, J. (1984) (2nd edn.), 'Phonologie: Intonation' (ch. 6), in K. E. Heidolph *et al.* (eds.), *Grundzüge einer deutschen Grammatik* (Berlin: Akademie-Verlag, 1st edn. 1980), 839–97.

PIERREHUMBERT, J. (1980), 'The Phonology and Phonetics of English Intonation', PhD dissertation (Massachusetts Institute of Technology).

PIKE, K. L. (1945), *The Intonation of American English* (Ann Arbor: University of Michigan Press).

REYELT, M., GRICE, M., BENZMÜLLER, R., MAYER, J., and BATLINER, A. (1996), 'Prosodische Etikettierung des Deutschen mit ToBI', in D. Gibbon (ed.), *Natural Language and Speech Technology, Results of the third KONVENS conference* (Berlin, New York: Mouton de Gruyter), 144–55.

SCHERER, K. R., LADD, D. R., and SILVERMAN, K. E. A. (1984), 'Vocal Cues to Speaker Affect: Testing Two Models', *Journal of the Acoustical Society of America*, 76/5: 1346–56.

SELTING, M. (1995), *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation* (Tübingen: Niemeyer).

SHATTUCK-HUFNAGEL, S., OSTENDORF, M., and ROSS, K. (1994), 'Stress Shift and Early Pitch Accent Placement in Lexical Items in American English', *Journal of Phonetics*, 22: 357–88.

TRAGER, G. L. and SMITH, H. L. (1951), *An Outline of English Structure* (Norman, Oklahoma: Battenburg Press, revised edition 1957, Washington, American Council of Learned Societies).

TRUCKENBRODT, H. (1998), 'On the Role of Prosody in Tonal Scaling', paper presented at Workshop on Focus, Intonation and Phrasing (Freudental, near Konstanz, Germany).

—— (2002), 'Upstep and Embedded Register Levels', *Phonology*, 19(1): 77–120.

UHMANN, S. (1991), *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie* (Tübingen: Niemeyer).

WUNDERLICH, D. (1988), 'Der Ton macht die Melodie—Zur Phonologie der Intonation des Deutschen', in H. Altmann (ed.), *Intonationsforschungen* (Tübingen: Niemeyer), 1–40.

4

Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora

Amalia Arvaniti and Mary Baltazani

4.1. INTRODUCTION

This chapter provides an analysis of the prosodic and intonational structure of Greek within the autosegmental/metrical framework of intonational phonology (Pierrehumbert 1980; Pierrehumbert and Beckman 1988; Ladd 1996), and presents Greek ToBI (henceforth GRToBI), a system for the annotation of Greek spoken corpora based on this analysis. Both the analysis and the annotation system have largely been developed on the basis of a corpus of spoken Greek, especially collected for this purpose, and including data from several speakers and a variety of styles (read text, news broadcasting, interviews, spontaneous speech). The linguistic variety analysed—and the one for which GRToBI was conceived and designed—is Standard Greek as spoken in Athens. It is our hope that other varieties of Greek will be similarly analysed, and that eventually GRToBI will be adapted for the annotation of corpora in those varieties as well.

4.2. STRESS AND RHYTHM

Greek is a stress accent language, described in traditional grammars as having 'dynamic stress' (among others, Joseph and Philippaki-Warburton 1987).

The development of GRToBI largely took place while the first author was on sabbatical leave from the University of Cyprus and visiting at the Ohio State University Linguistics Laboratory. We are grateful to the *Labbies*, particularly Mary Beckman, Julie McGory, Shu-hui Peng, Amanda Miller, and Mariapaola D'Imperio, for their support and input. Thanks are also due to our editor Sun-Ah Jun, to the students in Mary Beckman's and Julie McGory's 1999 ToBI course, to Bob Ladd, and particularly to Janet Fletcher whose review got us re-thinking. Last but not least, we would like to express our gratitude to Jenny and Peter Ladefoged for their kind hospitality to the first author during her stay in Los Angeles.

Stress is acoustically manifested as either longer duration or higher amplitude of the stressed syllable (or both), making *total amplitude* the most robust cue to stress (Arvaniti 1991, 2000).

Primary stress cannot be predicted from phonological structure, as there are no syllable weight distinctions in Greek and stress is not fixed. The main phonological limitation on the position of primary stress is that it falls on one of the last three syllables of the word (Joseph and Philippaki-Warburton 1987; Setatos 1974). In all other aspects, stress placement within this three-syllable window is largely determined by morphology (Drachman and Malikouti-Drachman 1999; Revithiadou 1998). At the lexical level, there are no other stresses in addition to primary stress.

The presence of *postlexical* rhythmic stresses, on the other hand, is disputed. Malikouti-Drachman and Drachman (1981), and Nespor and Vogel (1989) suggest that rhythmic stresses appear regularly to remedy stress lapses in Greek. This claim, however, is not supported either by acoustic evidence or by the native speakers' intuition (see Arvaniti 1994, and references therein). Thus, we assume here that at the postlexical level, as well as lexically, there is only one stressed syllable per word.

There is one regular exception to this stipulation, however: content words stressed on the antepenult (or the penult) and followed by one (or two) enclitics acquire an additional stress two syllables to the right of their lexical stress; e.g.

(1) (*a*) /ˈfernodas to mu/>[ˌfernoˈdastomu]
bringing it to-me
(*b*) /to teˈtraðio mu/>[to teˌtraðiˈomu]
the notebook my

For convenience, we will follow the practice of Greek grammarians and call the added stress of such sequences 'enclitic stress', although it does not usually fall on the enclitic itself, as examples (1*a*) and (1*b*) show.

In contrast to words with enclitic stress, there are several disyllabic function words in Greek (e.g. (2*a*)–(2*c*)) which are normally uttered without stress, and thus form part of the following content word (with which they are syntactically, as well as prosodically, linked). Some of these words contrast with stressed homophones (cf. (2*c*) and (2*d*)).

(2) (*a*) /aˈpo noˈris/>[aponoˈris]
since early
(*b*) /aˈna tin iˈfilio/>[anatiniˈfilio]
all-over the globe
'everywhere'

(*c*) /ka'ta to 'spiti/>[katato'spiti]
towards the house
(*d*) /ka'ta tu 'ɣamu/>[ka'ta tu'ɣamu]
against (the) marriage

With respect to rhythm, Greek has been described as syllable-timed, though Dauer (1983) places it somewhere in the middle of the continuum between stress- and syllable-timing. Arvaniti (1994), however, points out that by Dauer's own criteria (1983, 1987) Greek should be a prototypical syllable-timed language, virtually devoid of stress; such a description, however, is not supported either by acoustic data (e.g. Arvaniti 1991, 1992, 2000; Botinis 1989), or by the fact that Greek has several minimal pairs and triplets distinguished solely by stress placement. Instead, Arvaniti (1994) proposes that rhythm in Greek (as in all languages) is based on the alternation of strong and weak prosodic constituents. The difference between languages described as stress-timed, e.g. English, and languages described as syllable-timed, e.g. Greek, lies in the fact that the former tend to keep this alternation as even as possible, by using various strategies to eliminate stress clashes and lapses; in contrast, the latter languages seem to allow for less eurhythmic patterns, i.e. they tolerate clashes and lapses to a greater extent.

4.3. INTONATIONAL PHONOLOGY

For the intonational analysis of Greek we recognize three types of tonal events: *pitch accents*, which associate with stressed syllables, and two types of phrasal tones, *phrase accents* and *boundary tones*, which associate with the boundaries of intermediate and intonational phrases, respectively. In contrast to stress, which as mentioned is lexically determined, the tones are morphemes that encode pragmatic information. Therefore, it is not expected that every stressed syllable will be accented (see also Section 4.4.1).

4.3.1. *The pitch accents*

Greek has five pitch accents, L*+H, L+H*, H*+L, H* and L*. By far the most frequently used pitch accent is L*+H, which is the predominant choice for pre-nuclear accented syllables. Because of its distribution, L*+H has been described as the 'pre-nuclear' accent of Greek (Arvaniti *et al.* 1998; Baltazani and Jun 1999). In our corpus, however, this accent was frequently attested in nuclear position, in calls, imperatives, negative declaratives, and wh-questions.

Phonetically, the L*+H is manifested as a gradual rise from a trough (the L tone) to a peak (the H tone). In canonical conditions, that is, if there are at least two unstressed syllables between consecutive L*+H accents, the L is aligned at the very beginning or slightly before the onset of the accented syllable, and the H early in the first post-accentual vowel (Arvaniti and Ladd 1995; Arvaniti *et al.* 1998). The rather atypical alignment of the tones in the L*+H accent has given rise to a great deal of fluctuation in its description (see also Arvaniti *et al.* 2000, for a discussion of the problems that the alignment of L*+H poses for the notion of starredness in intonational phonology). In GRToBI this accent is analysed as L*+H, because our corpus and other quantitative data (Arvaniti *et al.* 1998; Arvaniti *et al.* in press) show that it is in contrast with another accent which can be most plausibly described as L+H*.

Specifically, as illustrated in Figure 4.1, in L+H* the H tone appears roughly in the middle of the accented vowel, unlike L*+H which shows late alignment of the H tone. L+H* is often used to signal narrow focus, as in Figure 4.9 (Arvaniti *et al.* in press; Baltazani in press; Baltazani and Jun 1999; Botinis 1998). (In Figure 4.1 and subsequent figures, the tonal label is shown on the first tier, followed by a phonetic transcription of each prosodic word in ASCII. The third and fourth tiers are the romanization of the words and the break indices, while the fifth tier is used to encode miscellaneous information that may be useful in interpreting the other tiers.)

The two bitonal accents are in contrast with the monotonal H* accent. As can be seen in Figure 4.2, H* lacks the initial dip associated with the L tone of L+H* and L*+H. Rather, there is a declining plateau between the H tone of the L*+H accent and the nuclear H* (hence the use of the downstep diacritic, !, which however *does not* denote a distinct phonological category; see Sections 4.3.2 and 4.5.1 for discussion). In Greek, this plateau does not exhibit the 'sagging' posited by Pierrehumbert (1981) as a possible type of interpolation between successive H*s in American English. When H* is used as the nuclear accent in a declarative utterance, as in Figure 4.2, it signals broad focus, and thus contrasts with L+H*, which signals narrow focus in the same context.

In nuclear position in declaratives, the H* also contrasts with H*+L. From a pragmatic point of view, H*+L conveys a more nonchalant (or even wearied) attitude on the part of the speaker than H*. Phonetically, it is realized as a fall from high pitch, with the fall being completed by the end of the accented syllable. It is useful to compare on this point the low stretch of the two contours shown in Figure 4.3, which illustrates the difference between H*+L and H*: as can be seen, when the pitch accent is H*+L the bottom of the speaker's voice is reached at the end of the accented syllable; in contrast, when the pitch accent is H*, the lowest Fo point is reached at the end of the first postnuclear syllable and the fall is realized in three discernible steps.

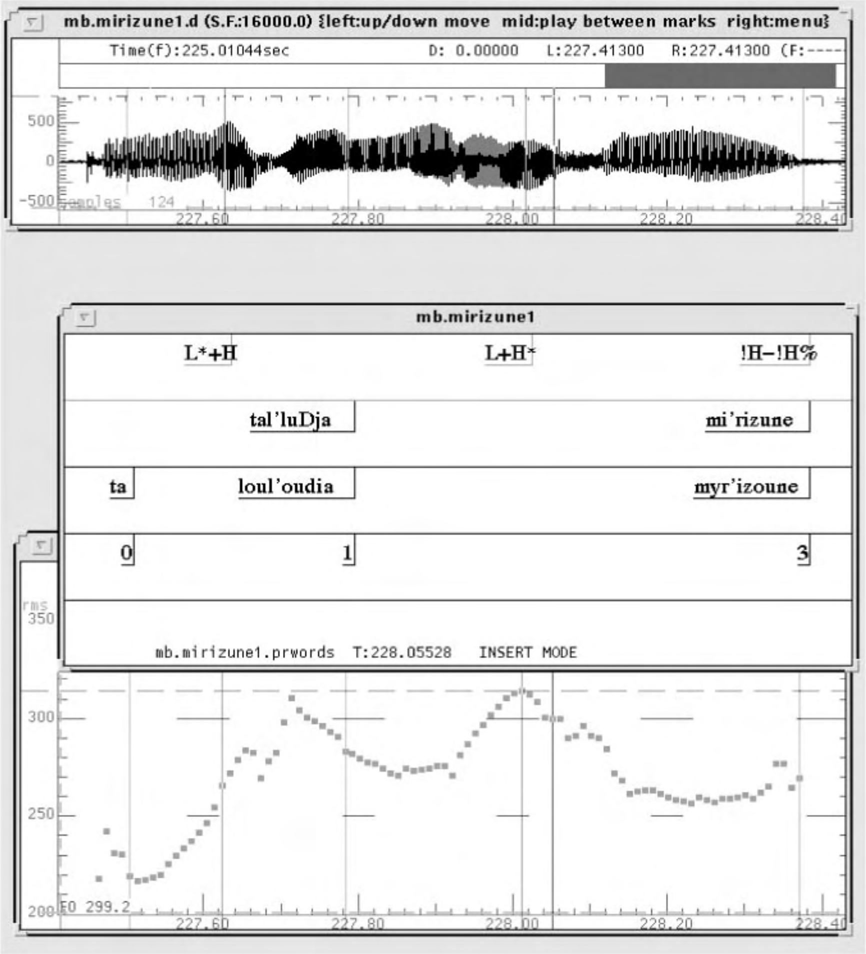


FIGURE 4.1 This example (the flowers smell 'Do the flowers really smell?') illustrates the different alignment of the H tone in L*+H (the accent on [l̩ˈluðʝa]) and L+H* (the accent on [miˈrizune]). In particular it shows how the H tone is aligned with the first postaccentual vowel in the former accent, but in the middle of the accented vowel in the latter. Further this figure shows a !H-!H% phrasal configuration, realized as a plateau in the middle of the speaker's range.

It should further be noted that in an earlier version of GRToBI we had analysed H*+L as !H* (Arvaniti and Baltazani 2000). We have now revised our position, to bring the use of the downstep feature more in line with that in other ToBI systems, in which downstep—for obvious reasons—cannot be used utterance-initially (as we would be forced to do in Figure 4.3). Furthermore, it is

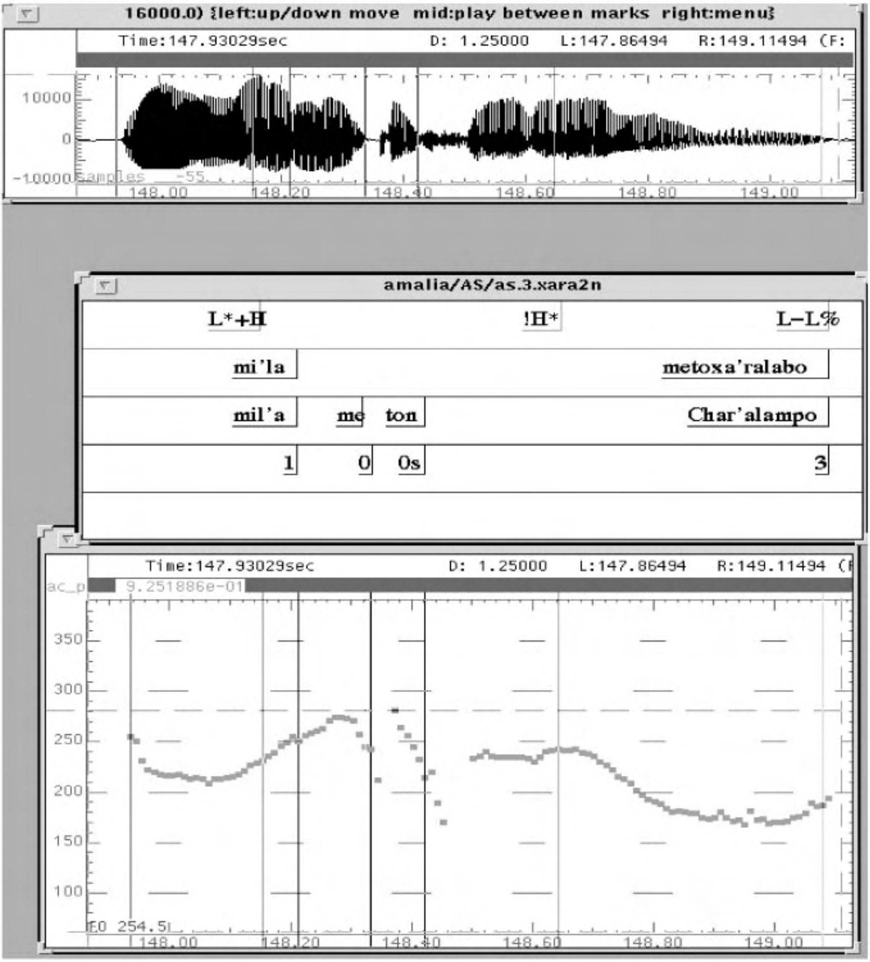


FIGURE 4.2 This example ('S/he is talking to Charalambos') shows a typical H* nuclear accent (downstepped, in this case). Note the lack of a dip at the beginning of this accent; cf. the L+H* in Figure 4.1.

clear that H* and H*+L are distinct both in terms of meaning and in terms of realization, and thus analysing H*+L as a 'scaled down' version of H* is not well motivated. Finally, it appears that !H* is needed for cases of 'scaled down' H* pitch accents, like the one shown in Figures 4.2 and 4.8. Despite these indications against our original analysis, it is not the case that the analysis of this accent as H*+L is entirely unproblematic; for example, it is not obvious why in Figure 4.3, H*+L is scaled lower than the utterance-beginning. Clearly then, the

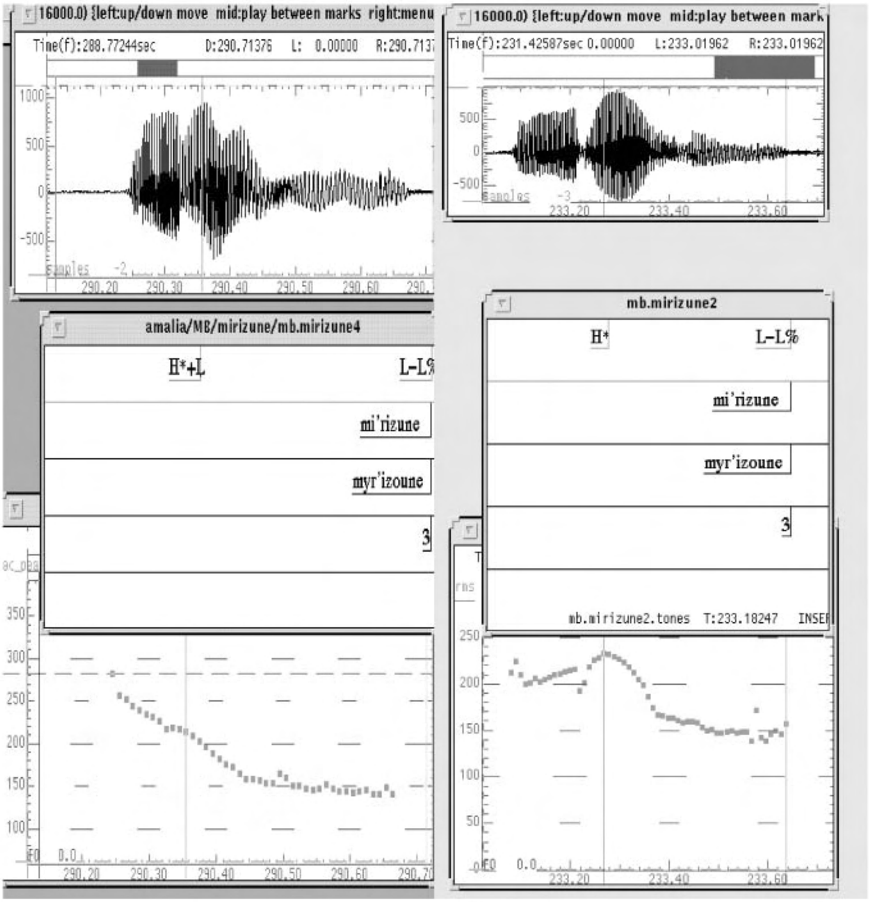

16000.0) {left:up/down move mid:play between marks right:menu
Time(f):288.77244sec D:290.71376 L: 0.00000 R:290.713
1000
500
0
-500
samples -2
290.20 290.30 290.40 290.50 290.60 290.
16000.0) {left:up/down move mid:play between mark
Time(f):231.42587sec 0.00000 L:233.01962 R:233.01962
500
0
-500
samples -3
233.20 233.40 233.60
amalia/MB/mirizune/mb.mirizune4
H*+L
L-L%
mi'rizune
myr'izoune
3
350
300
250
200
150
100
F0 0.0
290.20 290.30 290.40 290.50 290.60 290.70
mb.mirizune2
H*
L-L%
mi'rizune
myr'izoune
3
mb.mirizune2.tones T:233.18247 INSE
rms
250
200
150
100
50
0 F0 0.0
233.20 233.40 233.60

Figure 4.3 These two examples (both 'they smell') illustrate the difference between H*+L (on the left) and H* (on the right) on a one-word utterance. As can be seen, the H*+L is falling throughout the accented syllable [ri], while the H* accent involves a shallow rise on the accented syllable.

utterances in Figure 4.3 and similar examples do not provide sufficient evidence on which to choose one alternative over the other. Rather, controlled experiments are needed to provide a definitive answer to this problem.

Finally, the L* accent is typically realized as a low plateau, as shown in Figure 4.4. The L* appears as the nuclear accent before a 'continuation rise' (Baltazani and Jun 1999), in yes-no questions (Arvaniti *et al.* 2006; Arvaniti *et al.* in press; Baltazani and Jun 1999; Baltazani to appear) and in the calling contour we term 'suspicious' (see Table 4.1).

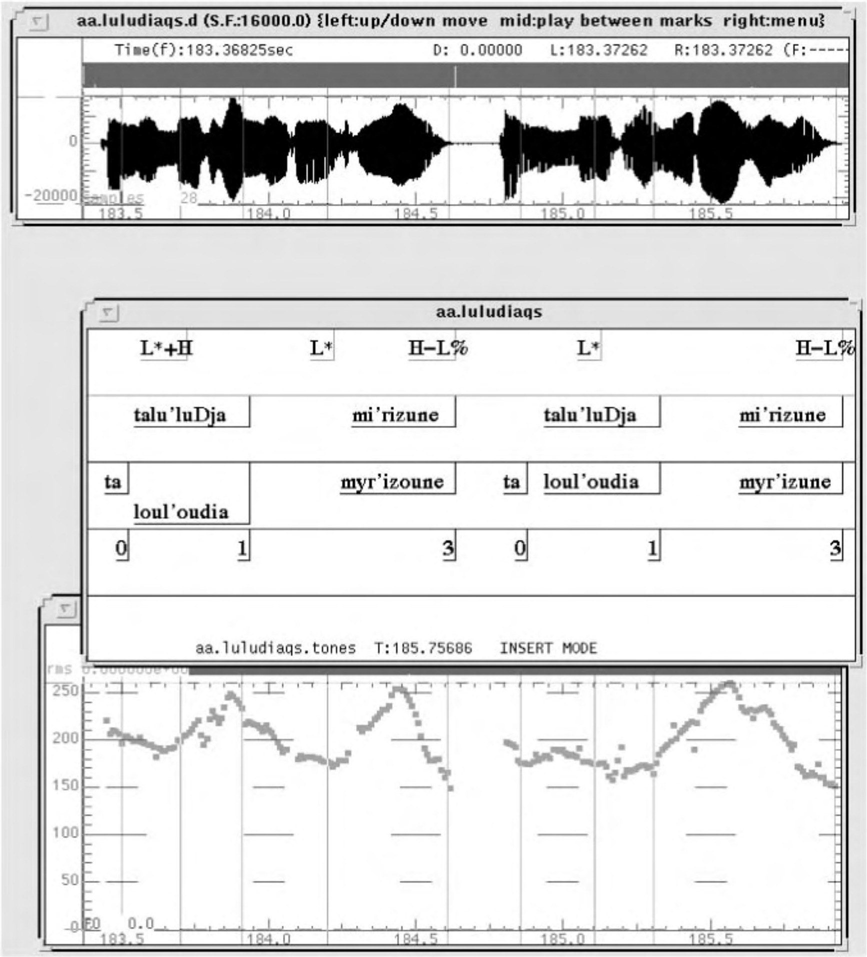


FIGURE 4.4 This illustration shows the same question twice (the flowers smell), with focus on the word [miˈrizune] 'they smell' on the left, and on the word [luˈluðʝa] 'flowers' on the right ('Do the flowers SMELL?' and 'Is it the FLOWERS that smell?' respectively). In both cases, the L* nucleus is realized as a low plateau with an additional dip in Fo on the accented syllable itself. Note also the different alignment of the H- in the two contours: as described in Section 4.3.4, the H- aligns with the unstressed penultimate syllable [zu] in the question on the left (where the nucleus is on the final word), but with the stressed syllable [ri] in the question on the right (where the nucleus is early).

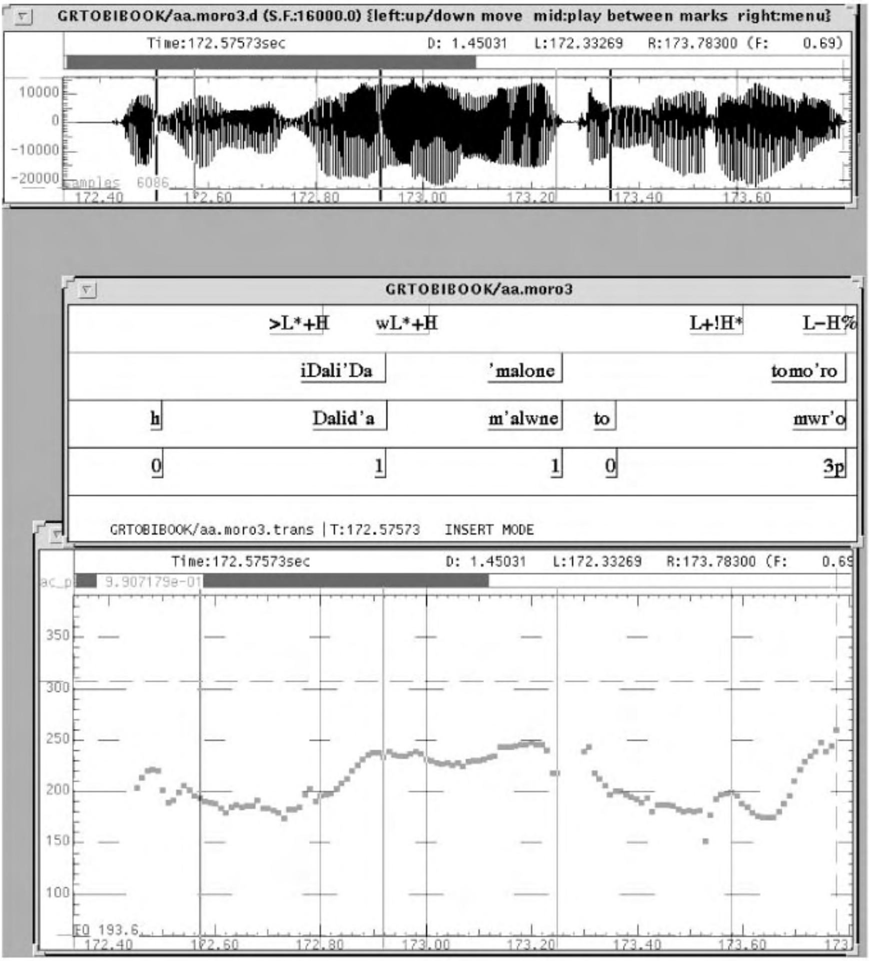

GRTOBIBOOK/aa.moro3.d (S.F.:16000.0) {left:up/down move mid:play between marks right:menu}
Time:172.57573sec D: 1.45031 L:172.33269 R:173.78300 (F: 0.69)
10000
0
-10000
-20000
samples 6086
172.40
172.60
172.80
173.00
173.20
173.40
173.60
GRTOBIBOOK/aa.moro3
>L*+H
wL*+H
L+!H*
L-H%
iDali'Da
'malone
tomo'ro
h
Dalid'a
m'alwne
to
mwr'o
0
1
1
0
3p
GRTOBIBOOK/aa.moro3.trans |T:172.57573 INSERT MODE
Time:172.57573sec D: 1.45031 L:172.33269 R:173.78300 (F: 0.69
9.907179e-01
350
300
250
200
150
100
F0 193.6
172.40
172.60
172.80
173.00
173.20
173.40
173.60
173

FIGURE 4.5 This example ('Dalida was scolding the baby [when she fainted]!') exemplifies the L-H% phrasal configuration, which is preceded in this case by a L+!H* accent on [mo'ro]. Note also the early aligned (>L*+H) and undershot (wL*+H) realizations of the two L*+H accents on the clashing [iðali'ða] and ['malone] respectively.

4.3.2. *Downstep and the phonetic realization of pitch accents*

The descriptions presented above provide a phonological analysis of the Greek accentual system and a sketch of the phonetics of the pitch accents under canonical conditions. However, pitch accents show significant contextual variability as regards both the scaling and the alignment of their targets, with tonal crowding and downstep being the main influences.

Concretely, L*+H and L* exhibit noticeable variability in contexts of tonal crowding, i.e. when several tones must be realized within a short segmental stretch. Previous research (Arvaniti 1994; Arvaniti *et al.* 1998, 2000) and the data of our own corpus show that the speakers adopt mainly three strategies to cope with the tonal crowding of consecutive L*+H accents. Specifically, they may (a) undershoot the L tone of the second L*+H accent; (b) realize the first accent earlier than normal and undershoot the second one (as in Figures 4.5 and 4.10); (c) realize the first accent earlier and the second one later than normal. Similarly, in cases of tonal crowding, L* accents are realized as rising from a low point, as in [moˈro] in Figure 4.10.

There are two complementary explanations for these resolutions of tonal crowding. First, Greek favours the undershooting of all underlying tones to the truncation of some of them (for the distinction between undershooting and truncating languages, see Grice *et al.* this volume Ch. 13; Ladd 1996). It follows that L*+H, which requires at least two syllables for its canonical alignment, will be the accent most prone to undershoot. Second, it appears that in Greek the undershooting of L tones is preferred to the undershooting of H*s*. Some support for this hypothesis comes from similar evidence on the undershooting of L% in Japanese (Venditti this volume Ch. 7), suggesting that different realization constraints may apply more generally to L and H targets (see also, Arvaniti and Garding, in press; Arvaniti *et al.* 2006; Arvaniti *et al.* in press). Regardless of the underlying reasons, the resolution of tonal crowding in Greek is still not entirely understood; e.g. it is not clear whether the strategies mentioned above are under the speaker's choice or depend on prosodic factors, such as phrasing and the relative metrical strength of the accented syllables.

To our knowledge, the role of downstep in Greek intonation has not been investigated before. Our data, however, allow us to make certain observations. First, it is clear that in Greek downstep is *not* triggered by the presence of bitonal accents. The hypothesis linking downstep to bitonal accents was first advanced in Pierrehumbert (1980) for English, and was further refined and extended in Beckman and Pierrehumbert (1986), who took both English and Japanese data into account. This hypothesis (which has its origins in studies of tonal phonology in African languages) is often taken to reflect a universal tendency of tonal implementation (e.g. Goldsmith 1999: 4). As mentioned, however, in Greek the most frequently attested pitch accent in prenuclear position is the bitonal L*+H. As most content words are accented in Greek, sentences with consecutive L*+H accents but no downstep, such as those in Figures 4.1 and 4.10, are quite common. This lack of scaling interaction among tones in Greek is further supported by similar data from the scaling of phrase accents and boundary tones, discussed at some length in Sections 4.3.3 and 4.3.4 (for additional evidence against the universality of the downstep trigger, see Yip 1996 and references therein).

In addition, our corpus suggests that certain scaling differences probably reflect phonetic regularities, and thus need not be part of the phonological description of Greek intonation. We refer, in particular, to the widespread lower scaling of the nuclear accent relative to previous accents, illustrated in Figures 4.2 and 4.5. One possible reason for this type of lower scaling could be *final lowering*, i.e. the progressive lowering of overall pitch range within the last stretch of an utterance (Liberman and Pierrehumbert 1984). However, evidence like that presented in Figure 4.5—in which the nuclear L+H* accent is clearly downstepped relative to the previous pitch accents, but the following H% is fully scaled—suggests that final lowering cannot be the only reason for the observed scaling effects. Although this type of downstep is regular and does not appear to have pragmatic significance—reasons for which we assume that it has no phonological bearing—it is clear that further research is necessary before final conclusions about the role and operation of downstep in Greek can be drawn.

4.3.3. *The phrase accents*

There are three types of phrase accent in Greek, H-, L- and !H-. The scaling of the H- and L- phrase accents does not appear to be influenced by the identity of neighbouring tones. This contrasts with the situation observed in other languages, such as English and German, in which the scaling of phrase accents is influenced by preceding or following tones, resulting in upsteps and downsteps (Pierrehumbert and Hirschberg 1990 and Beckman and Ayers-Elam 1997, on American English; Grice and Benzmüller 1995, on German). Because of this difference between Greek and other languages, falls or rises to mid pitch cannot be attributed to the upstepping or downstepping influence of a neighbouring tone and are thus represented by !H-.

4.3.4. *The boundary tones*

Greek has three types of boundary tone, H%, L% and !H%. As with the phrase accents, !H% is used to represent mid-level pitch, since phrase accents do not trigger the upstep or downstep of boundary tones. Note, for example, that both L-H% and L-!H% are attested in Greek; that is, L- cannot be seen as the trigger of the downstep of !H% in the latter case, since it does not affect the scaling of

H% in the former. Furthermore, no upstep of L% boundary tones due to a preceding H tone has been attested in the GRToBI corpus (see e.g. Figure 4.4).

The three boundary tones combine with the phrase accents in eight different configurations that appear to have specific pragmatic functions. The possible phrase accent-boundary tone combinations and their typical usage are shown in Table 4.1.

The L-L% configuration is manifested as a low plateau at the end of an utterance, illustrated in Figures 4.2, 4.3, 4.9 and 4.10. The L-L% is preceded by L+H*, H* or !H* in declaratives, by L+H* in imperatives and negative declaratives, and by L*+H in wh-questions (Arvaniti, Ladd and Mennen 1999; Arvaniti 2001).

The L-H% is manifested phonetically as a dip and then a rise to a high Fo value. The L-H% is often found after a L+H* pitch accent, with the whole configuration (L+H* L-H%) suggesting an 'involved' type of continuation rise, that raises the expectations of the hearer about what is going to follow. An example of this type of rise is shown in Figure 4.5. The L-H% is preceded by L* in the 'suspicious' calling contour mentioned in 4.3.1.

The H-L% configuration is used in yes-no questions, in which the nuclear accent is invariably L*. The H- accent in this case shows two distinct patterns of alignment, depending on the position of the nucleus (Baltazani and Jun 1999; Grice *et al.* 2000; Baltazani to appear; Arvaniti *et al.* 2006). Specifically, if the nucleus of the question is *not* on the final word of the utterance, the H-aligns with the stressed syllable of the final word. If the nucleus is on the final

TABLE 4.1 Possible combinations of phrase accent and boundary tone and their usage

Configuration	Schematic representation	Usage
L-L%	——	declaratives, negative declaratives, imperatives, wh-questions
L-H%	╲╱	'involved' continuation rise, 'suspicious' calls
H-L%	╱╲	yes-no questions, requesting calling contour
H-H%	╱	continuation rise, questioning calling contour
L-!H%	——╱	'involved' wh-questions, negative declaratives showing reservation, requesting imperatives
H-!H%	╱——	stylized continuation rise
!H-!H%	╲——	stylized call, incredulous questions
!H-H%	—╱	polite stylized call

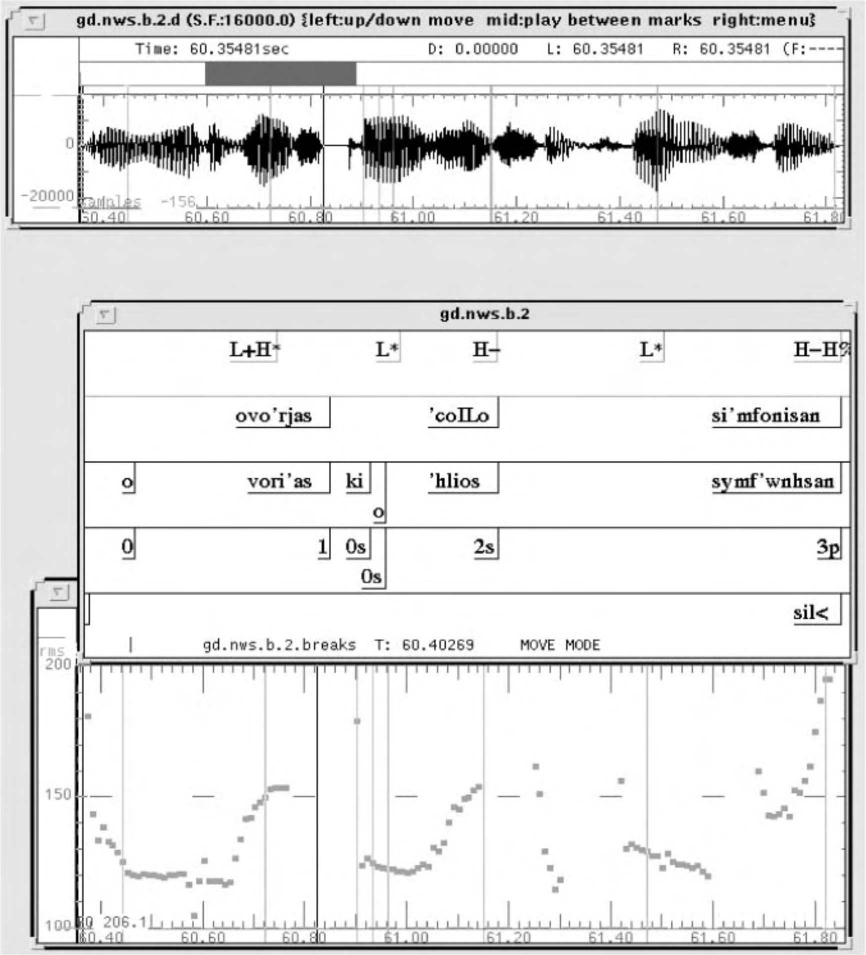

gd.nws.b.2.d (S.F.:16000.0) {left:up/down move mid:play between marks right:menu}
Time: 60.35481sec D: 0.00000 L: 60.35481 R: 60.35481 (F:----
0
-20000
samples -156
60.40
60.60
60.80
61.00
61.20
61.40
61.60
61.80
gd.nws.b.2
L+H*
L*
H-
L*
H-H%
ovo'rjas
'coILo
si'mfonisan
o
vori'as
ki
o
'hlios
symf'wnhsan
0
1
0s
0s
2s
3p
sil<
gd.nws.b.2.breaks T: 60.40269 MOVE MODE
rms
200
150
100
0 206.1

FIGURE 4.6 This example (‘The north wind and the sun agreed . . . ’) illustrates the much higher scaling of H-H% relative to H-. Further the example shows the diphthongization of /o/ and /i/—pronounced as a rising diphthong—in /ke o ˈiʎos/ ‘and the sun’, a phrase realized as one PrWd, [ˈcoi.ʎos], as evidenced by the alignment of the L* with the whole syllable [coi]. Finally, this example illustrates consonant degemination across ip boundaries (BI 2s): /ke o ˈiʎos_simˈfonisan/>[ˈcoiʎosiˈmfonisan].

word, the H- and L% are realized at the right edge of the utterance. These two alignment patterns can be clearly observed in Figure 4.4.

The H-H% configuration is manifested as a smooth rise to a high F0, as shown in Figure 4.6. It is typically preceded by a L* in both ‘continuation

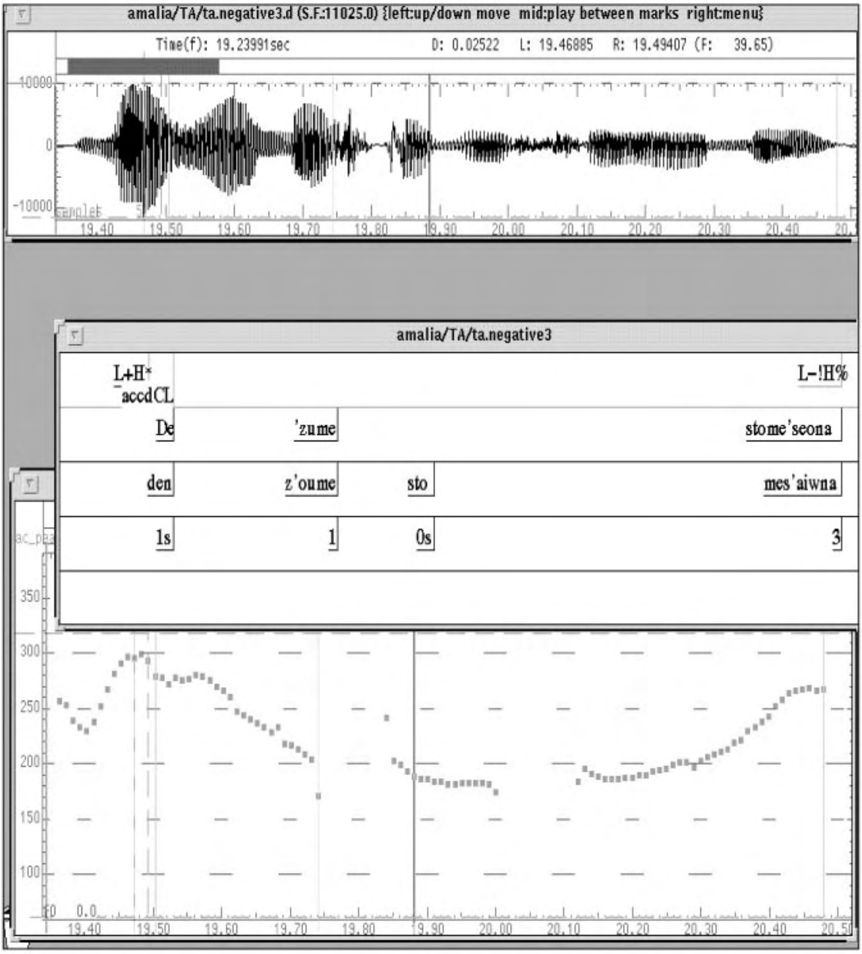


FIGURE 4.7 This example ('We do not live in the Middle Ages!') illustrates the typical pattern of a negative declarative expressing reservation. Note that the negative particle /ðen/, which is considered a phonological clitic, carries the nuclear (and only) pitch accent of the utterance, and thus forms a separate PrWd from the de-accented verb [ˈzume] 'we live'; yet, sandhi (/n/-deletion before the fricative [z]) does take place as well, although it is said to take place only within PrWd boundaries (Nespor and Vogel 1986). The rest of the utterance is deaccented, with the L- spreading until the last syllable ([na] of [meˈseona] 'Middle Ages'). Finally, compare the scaling of the !H% (relative to that of the L+H* peak) to the scaling of the H- and H% tones in Figures 4.5 and 4.6 (relative to the accentual H tones in those examples).

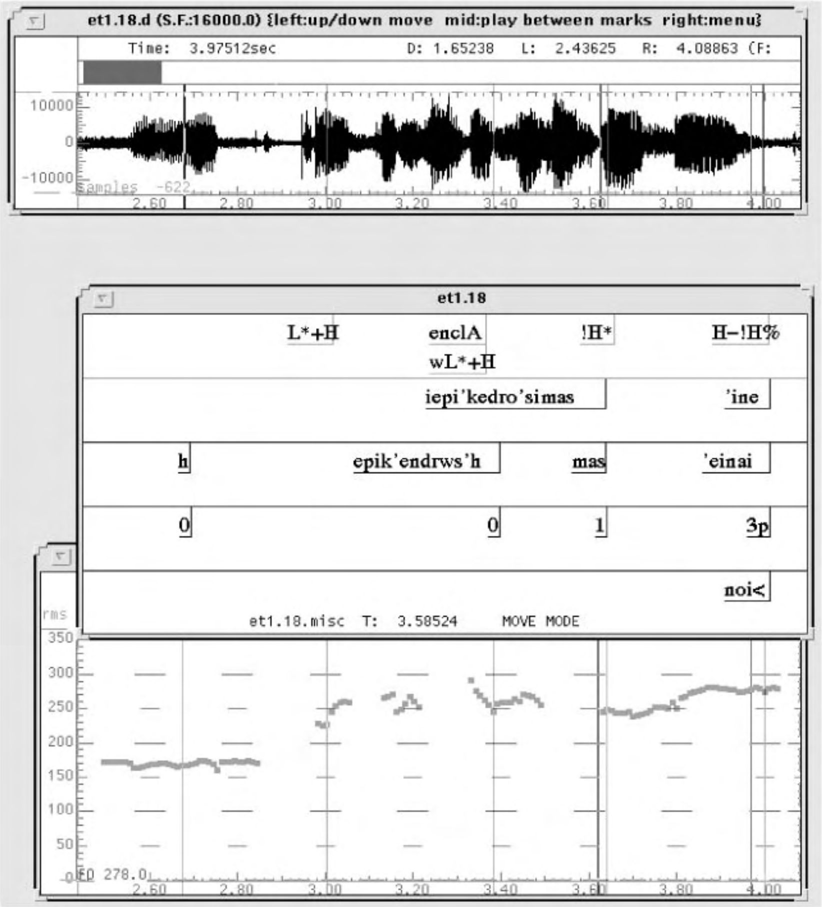


FIGURE 4.8 This example ('our focus is...') illustrates the stylized H-!H% configuration on the word ['ine] 'is'. Note also the presence of two accents on the word [epiˌcedro'si] 'focus', which here is followed by the enclitic [mas] 'ours', and thus carries enclitic stress on its last syllable [si].

rises' and in the questioning calling contour. The L-!H% is found in wh-questions, requesting imperatives, and negative declaratives that show reservation. All these types of utterance have similar intonational structure: a L*+H or L+H* nucleus followed by a low plateau (a spreading L-), and a small rise (the !H%). As can be seen in Figure 4.7 which illustrates this contour, the !H% remains approximately in the middle of the speaker's range.

The stylized configurations—H-!H%, !H-!H%, !H-H%—are used less often than the rest and for a limited number of pragmatic purposes. The

H-!H%, illustrated in Figure 4.8, is realized as a rise to a high plateau. Our corpus suggests that it is employed mostly when the speaker wants to hold the floor while preparing his/her next utterance, but s/he is unwilling to use a (filled) pause. The !H-!H% is the mirror image of H-!H%, that is, a fall to a plateau in the middle of the speaker's range; it is used in the *vocative chant* after L*+H, and also in incredulous questions after L+H*, as illustrated in Figure 4.1. Finally, the !H-H% is similar to !H-!H%. The difference between the two is the small rise at the end of the plateau in !H-H%, which makes the utterance sound more tentative or polite.

4.4. PROSODIC STRUCTURE

In the analysis adopted here, we assume that Greek has only three prosodic constituents at and above the word: the prosodic word, the intermediate phrase, and the intonational phrase. As we show below, there is ample stress, tonal, and sandhi evidence in support of these three prosodic levels.

4.4.1. *The prosodic word*

A prosodic word (henceforth PrWd) consists of a content word and its clitics. The term 'clitic', as used here, includes all items that in a given utterance lose their stress and form one PrWd with a host. As mentioned in Section 4.2, in Greek this is the common fate of most function words, including disyllabic ones (which are *not* traditionally considered to be phonological clitics).

A PrWd is expected to have only one stress; consequently it may bear only one pitch accent, though it is also possible for a PrWd not to be accented at all; this is, for example, the case with postnuclear PrWds (see Figures 4.7 and 4.9). PrWds with enclitic stress, on the other hand, may have two pitch accents, one on the lexically stressed syllable of the host and one on the syllable with enclitic stress (PrWds with enclitic stress may of course be de-accented; however, if they have only *one* pitch accent, this will necessarily fall on the syllable with enclitic stress).

In addition to stress and tonal cues, PrWds are the domain of at least eight types of sandhi, exemplified below in (3–10). Some of these—stop-voicing, /n/-deletion, /s/-voicing, vowel degemination, and vowel deletion—have been reported in Kaisse (1985), Nespor and Vogel (1986), and Condoravdi (1990), though their descriptions do not always match our data (see Section 4.4.3). The other three rules emerged from our corpus.

- Stop-voicing after a word-final nasal (the nasal is usually deleted; if not, it assimilates to the place of articulation of the stop); e.g.

(3) /tin 'poli/>[ti'boli] or [ti'mboli]
the town ACC.

- /n/-deletion before sonorants and fricatives; e.g.

(4) /ton la'o/>[tola'o]
the people ACC.

- /s/-voicing before sonorants; e.g.

(5) /o'ʝos mu/>[o'ʝozmu]
'my son'

- vowel degemination; e.g.

(6) /ta 'atoma/>['tatoma]
the individuals

- deletion of one of non-identical vowels; e.g.

(7) /to 'atomo/>['tatomo]
the individual

- /n/-resyllabification before a word-initial vowel (in accented syllables /n/-resyllabification is evident from tonal alignment); e.g.

(8) /'o.tan. 'e.fta.se/>[o.ta.'ne.fta.se]
when s/he arrived

- consonant degemination; e.g.

(9) /o'ʝos su/>[o'ʝosu]
'your son'

- diphthongization of non-identical vowels; e.g.

(10) /o.'i.ʎos/>['oi.ʎos]
the sun

4.4.2. *The intermediate phrase and the intonational phrase*

The two levels of phrasing above the PrWd are the intermediate and the intonational phrase (ip and IP respectively). An ip must include at least one pitch accent (i.e. there are no headless phrases in Greek), and is tonally demarcated by the presence of a phrase accent (H-, L- or !H-) at its right edge. An IP must include at least one ip, and is tonally demarcated by the presence of a boundary tone (H%, L% or !H%) at its right edge.

There is abundant evidence for these two levels of phrasing in Greek. First, the tones associated with ip*s* show a simple Fo movement, such as a fall or a rise, unlike the right edges of IP*s* which often show more complex pitch configurations (see Section 4.3.4). In the cases where the pitch movement is of the same type (e.g. a rise), ips and IPs show a difference in

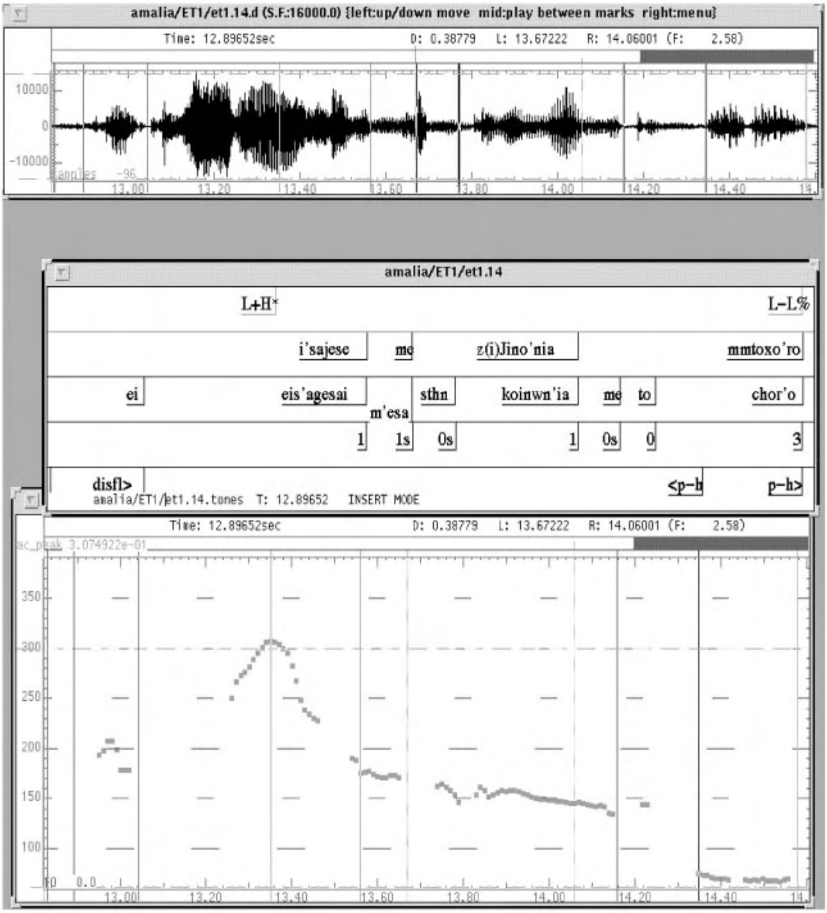


FIGURE 4.9 This example ('You BECOME-PART of society through dance') illustrates de-accenting after early focus. Note also, the several instances of sandhi and fast speech rules. The phonological representation of this utterance is /iˈsaɣese ˈmesa stin kinoˈnia me to xoˈro/ and the expected realization, according to phonological descriptions, is [iˈsaʝese ˈmesa stiɲɟinoˈnia metoxoˈro]; the actual realization is [iˈsaʝese ˈmeziɟinoˈnia m̩toxoˈro]. Note finally the presence of pitch-halving at the end of the utterance and the use of the pitch-halving label in the Miscellaneous Tier.

scaling, as illustrated in Figure 4.6. This observation is supported by quantitative data: after the level of phrasing had been assigned (independently of scaling) and agreed upon by the authors, a systematic comparison was made of the difference in Hz between the lowest and highest Fo point in L* H- and L* H-H% configurations in the data of four speakers reading *The North Wind and the Sun*. The results showed that this difference was smaller in L* H- than in L* H-H%, i.e. the peak was scaled lower in the former case by 30 Hz on average.

On the other hand, in IP*s* with complex final movement, such as a rise-fall, the two tones align independently (as noted, for example, in the description of H-L% and L-!H% in Section 4.3.4 and illustrated in Figures 4.4 and 4.7 respectively). This clearly shows that we are not dealing with bitonal boundary tones; if that were the case, then the individual tones would be expected to align together at the edge of the relevant phrase. Rather, in the L-!H% melody the L- spreads, while the !H% aligns with the last vowel of the utterance (Arvaniti *et al.* 1999; Arvaniti 2001); in the H-L% melody the H- aligns with a stressed vowel, if one is available, while the L% always aligns with the last vowel (Arvaniti *et al.* 2006).

Grice *et al.* (2000) reviewed these Greek data (as well as related data from German, Hungarian, Romanian, and English), and concluded that this behaviour of phrase accents can be accounted for if we view phrase accents as phrasal tones with a secondary association to a specific tone bearing unit. This position does not of course account for the fact that phrase accents, in Greek at least, appear to always align at the edge of a non-final ip, and move to their secondary association site only when there is a boundary tone following. One possible reason for this difference between non-final and final ip*s* is that in the former the delimitative function of the phrase accent takes priority, while in the latter, this function is assumed by the boundary tone and thus need not be fulfilled by the phrase accent itself (for a detailed analysis along these lines, see Grice and Truckenbrodt 2001).

In addition to the tonal evidence, our corpus suggests that at least some types of sandhi take place within ip boundaries but not across them. One such case of sandhi relates to vowel hiatus, which is resolved within ip boundaries but not across them (an observation confirmed by Arvaniti and Pelekanou 2002, and Baltazani 2006*a*). On the other hand, Figure 4.6 illustrates another type of sandhi, consonant degemination, which does apply across an ip boundary (though not across IP boundaries). Finally, evidence for the two levels of phrasing comes from pauses: IP*s*, even non-final ones, may be followed by a lengthy pause, while pauses are rare after ip*s* and always very short.

4.4.3. *Sandhi and prosodic phrasing*

Although in the prosodic analysis presented here we assume only three constituent levels, in previous analyses of the prosodic structure of Greek, additional levels have been posited. Concretely, Nespor and Vogel (1986) propose that Greek prosodic structure includes the clitic group—the need for which has been generally disputed (e.g. Zec and Inkelas 1991)—and three phrasal constituents, the phonological phrase, the intonational phrase, and the phonological utterance. An additional constituent, the minimal phrase, was later proposed by Condoravdi (1990).

Although a full discussion of these analyses is beyond the scope of this work, it should be noted that evidence for these additional constituents comes exclusively from sandhi. However, many of the sandhi phenomena used to support these analyses have not been reliably described and analysed, resulting in disagreements between the phonological descriptions (such as Kaisse 1985; and Nespor and Vogel 1986) on the one hand, and naturally occurring data (such as those of Fallon 1994) on the other.

The examination of our own corpus allows us to make the following observations regarding sandhi. First, several types of sandhi apply across larger constituents than has previously been suggested. The sequence [ˈotaˈxtips] in Figure 4.10 is a case in point: the adverb /ˈotan/ 'when' loses its final /n/ before the verb /ˈxtipise/ 'rang', although /ˈotan/ and /ˈxtipise/ form separate PrWds (e.g. both remain stressed). According to Nespor and Vogel (1986) however, /n/-deletion before fricatives applies only within PrWd boundaries (to be precise, within the Clitic Group, which corresponds to our PrWd). This example is not an isolated instance, and cannot be attributed to fast speech, as there is evidence that the utterance it is part of (and which was elicited under laboratory conditions) was rather carefully enunciated; this evidence comes from the words /ˈmalone/ and /tiˈlefono/ which are realized as such, rather than as [ˈmal̩ne] and [ˈtlefono] respectively, as would be expected in fast casual speech.

Second, the application of some rules presented in Kaisse (1985) and Nespor and Vogel (1986) depends on the lexical items used, something rather unusual for postlexical rules (see, for example, Arvaniti 1991; and Malikouti-Drachman and Drachman 1992, for a discussion on the above-mentioned rule of /n/-deletion). Third, sandhi does not appear to be obligatory at any level, as Nespor and Vogel suggest about certain rules; the speaker may choose to apply a particular rule, or she may not. Finally, it appears that at least some of the rules involve gradient, rather than categorical, changes. This observation was supported by Arvaniti and Pelekanou (2002) and Baltazani (2006*b*), who show that gradience holds particularly true of /s/-voicing and vowel-deletion (both described in Section 4.4.1): in many instances of /s/-voicing, the /s/ is only partially voiced, while complete deletion of a vowel under hiatus appears to be very rare; in most cases, audible and (spectrographically) visible evidence of the 'deleted' vowel remains in the signal. These findings are not surprising, as they agree entirely with results reported in studies of similar phenomena in English and other languages (among several, Holst and Nolan 1995; Zsiga 1997; Ellis and Hardcastle 1999). Nevertheless, they strongly suggest the necessity of empirically re-examining the phonological descriptions of Greek sandhi in particular, and of the reliability of sandhi as a phrasing marker in general.

4.5. THE GRToBI ANNOTATION SYSTEM

As mentioned, GRToBI is a system for the prosodic annotation of Greek spoken corpora. The system is based on the Mainstream American English ToBI (Silverman *et al.* 1992; Beckman *et al.* this volume Ch. 2), but it has been adapted so as to take into account additional facets of Greek prosody (such as extensive sandhi) that merit particular attention. GRToBI has five tiers. The *Tone Tier*, that gives the intonational analysis of the utterances; the *Prosodic Words Tier*, which is a fairly narrow phonetic transcription; the *Words Tier*, that gives the text in romanization; the *Break Index Tier*, showing indices of cohesion; and finally the *Miscellaneous Tier*, in which other information may be entered. Details on each tier are given below (for labelling conventions see Appendix I, and for a summary of GRToBI labels see Appendix IV).

4.5.1. *The Tone Tier*

As mentioned, the Tone Tier presents the intonational structure of the utterance, using the analysis and criteria presented in Section 4.3. In addition to the pitch accents, phrase accents, and boundary tones described in that section, some diacritics are also used in the GRToBI annotation system. These are largely employed to provide a more detailed description of the phonetic realization of the pitch accents, in order to shed light on the relation between the phonological representation of accents and their context-dependent phonetics.

Concretely, L*+H pitch accents in tonal crowding contexts, in which as mentioned earlier their realization varies, are annotated using three diacritics: wL*+H ('w' for *weak*) is used when the L tone is undershot, as in ['malone] in Figure 4.10; >L*+H is used when the accent is realized *earlier* than typically expected, as in [ðali'ða] in the same figure; and <L*+H is used when the accent is realized later than typically expected. Similarly, undershot L*

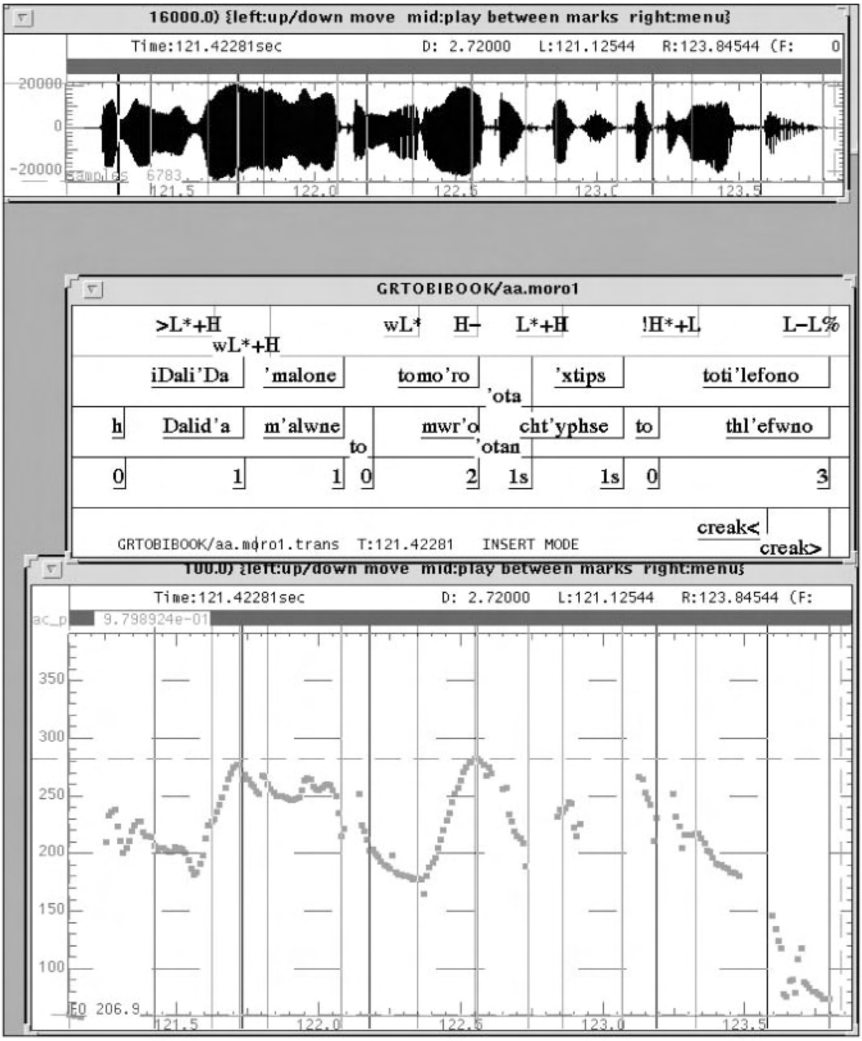

16000.0) {left:up/down move mid:play between marks right:menu}
Time:121.42281sec D: 2.72000 L:121.12544 R:123.84544 (F: 0
20000
0
-20000
samples 6783
121.5
122.0
122.5
123.0
123.5
GRTOBIBOOK/aa.moro1
>L*+H
wL*+H
wL*
H-
L*+H
!H*+L
L-L%
iDali'Da
'malone
tomo'ro
'ota
'xtips
toti'lefono
h
Dalid'a
m'alwne
to
mwr'o
'otan
cht'yphse
to
thl'efwno
0
1
1
0
2
1s
1s
0
3
creak<
creak>
GRTOBIBOOK/aa.moro1.trans T:121.42281 INSERT MODE
100.0) {left:up/down move mid:play between marks right:menu}
Time:121.42281sec D: 2.72000 L:121.12544 R:123.84544 (F:
ac_p 9.798924e-01
350
300
250
200
150
100
F0 206.9
121.5
122.0
122.5
123.0
123.5

FIGURE 4.10 This example ('Dalida was scolding the baby when the phone rang') shows two different realizations of L*+H under tonal crowding, >L*+H, which is realized earlier than it canonically would (the H tone is aligned with the accented vowel, instead of the first postaccentual vowel), and wL*+H, in which the L* tone is undershot, while the H shows the typical late alignment of H in L*+H accents. In this utterance there is also an undershot L* (wL*) on [mo'ro], realized as a rise from low pitch throughout the accented syllable (cf. the canonical L*s in Figure 4.6).

accents, usually realized as a low *point* rather than a plateau, are annotated as wL* (see Figure 4.10).

Further, the downstep diacritic (!), may be used with any of the pitch accents with a H tone, if the transcriber feels that the accent is scaled lower than declination warrants. As mentioned in Section 4.3.2, it is not clear what the role of downstep is in Greek. For this reason, we have decided to explicitly annotate downstep in the Tone Tier, even in cases in which we have reason to believe that the presence of downstep is phonetically determined (as is probably the case with the scaling of the nuclear accents in Figures 4.2 and 4.10). Explicit marking of downstep will facilitate further research, which can illuminate the scope and function of downstep.

4.5.2. *The Prosodic Words Tier*

The Prosodic Words Tier provides a detailed phonetic transcription of the utterances. Currently ASCII characters (with a fairly transparent relation to their IPA equivalents) are employed. We hope that in the future the information on this tier will be presented in IPA notation (for the current conventions see Appendix II).

In this tier, each PrWd constitutes one label. The aim of the PrWords Tier is to provide the users of the database with information about the actual pronunciation of the utterances. To this purpose the transcription is phonetic rather than phonological, that is, it encodes stress, allophonic variation, phone deletions, assimilations, and sandhi.

This tier was deemed necessary for two reasons. First, it facilitates the analysis of sandhi and fast speech rules, which abound in Greek, by encoding their outcome. Second, it provides information about stress. This information cannot be deduced from the transliteration or from Greek spelling conventions, since Greek orthography marks stress only on polysyllabic words. In a given utterance, however, a monosyllabic content word will most likely be stressed and accented, while a disyllabic function word will most often be cliticized. By coding and examining such cases we hope that a better understanding of the relation between stress and accent in Greek will emerge.

4.5.3. *The Words Tier*

At present the Words Tier provides a word-by-word romanization of the text, although our long-term goal is to present this information in Greek orthography. In the absence of a generally agreed system for the romanization of Greek, we have followed some of the more widely accepted conventions (such as *ch* for χ), and have devised means for transliterating the rest of the characters as transparently as possible. Our aim has been to represent each Greek letter and combination of letters with a unique roman character or set of characters, so that (a) searches of the Words Tier in the database yield unambiguous results and (b) the future algorithmic conversion to the Greek alphabet is possible. The full set of transliteration conventions can be found in Appendix III.

4.5.4. *The Break Index Tier*

GRToBI uses four levels of break indices, 0, 1, 2, and 3. These levels correspond to a *subjective* sense of increasing disjuncture between words. By *word* here we mean any item that is separated by spaces in the orthography of Greek; *orthographic* words often form but part of a *prosodic* word. It should also be stressed that although the use of a particular index relies on the transcriber's judgement, indices correlate with specific stress, sandhi, and tonal events, which the transcriber must take into consideration before reaching a decision.

BI 0 is used to mark boundaries within a sequence of orthographic words that show total cohesion of the type typically expected between items that form one PrWd. Thus, we assume that a sequence of orthographic words separated by BI 0 corresponds to a PrWd that has only one stressed syllable and may bear only one pitch accent. As noted, cases with two accents due to enclitic stress are also felt to form one PrWd. Because of this sense of cohesion, the boundaries between hosts and enclitics are labelled BI 0. However, little is as yet known of the intonational behaviour of such sequences (but see Arvaniti 1992; Botinis 1998). Since this is still an open research question, we decided to flag the second accent in these cases by adding a label to it, namely 'enclA' (for *enclitic accent*), as shown in Figure 4.8.

Although, as noted, several types of sandhi take place across a BI 0 boundary, its presence is not a necessary condition for BI 0 to be used (as it is in MAE-ToBI for instance; Beckman and Ayers-Elam 1997). For example, several forms of the Greek verbs include the proclitic particles, /θa/ or /na/; when the following verb stem begins with a consonant, no sandhi takes place between the particle and the verb. However, native speakers feel that these particles cannot be conceived but as part of the verb form; for this reason, BI 0 is marked in such cases.

BI 1 marks boundaries between PrWds. The presence of an accent should be considered crucial for deciding that an item is a distinct PrWd. Thus, when articles (which are normally proclitics) are accented—as often happens in media-speech (Arvaniti 1997)—then they are separated by BI 1 from the nouns that would normally be their hosts; they may also be flagged with 'accdCL' (for *accented clitic*) in the Tone Tier. On the other hand, it should be stressed that the absence of accent does not constitute evidence that a given stretch is *not* a PrWd; for instance, in cases of early focus in an utterance, de-accenting of all PrWds following the nucleus is expected (Arvaniti *et al.* in press; Baltazani in press; Baltazani to appear; Baltazani and Jun 1999; Botinis 1998), as illustrated in Figures 4.7 and 4.9.

BIs 2 and 3 mark ip*s* and IP*s* respectively.[1] The arguments for these two levels of phrasing and a description of the tonal and other prosodic cues that accompany each of them are presented in detail in Section 4.4.2.

In addition to the break indices, four diacritics are used to provide more detail on the prosodic structure of the annotated utterances: 's' for *sandhi*, 'm' for *mismatch*, 'p' for *pause* and '?' for *uncertainty*.

The diacritic *s* is used to flag *all* instances of sandhi at all prosodic levels, independently of whether sandhi rules operating at this level have previously been described for Greek or not. We hope that by investigating a large corpus of spoken data thus marked, a better understanding of the relation between sandhi, phrasing, and prosodic structure will be reached. Already the acoustic investigation of instances of sandhi in the GRToBI corpus (Arvaniti and Pelekanou 2002) and further experimental evidence (Baltazani 2006*a*, 2006*b*) have largely confirmed the more informal observations presented in Section 4.4.3, and have provided a preliminary distinction between categorical postlexical rules and rules of gradient phonetic implementation.

The *m* diacritic flags two types of mismatch between a given break index and the prosodic or tonal cues for it. The *m* diacritic is used with BI 0 to mark cases in which the context for sandhi at BI 0 exists, but sandhi does not take place. For example, when a sequence such as /tin ˈkori/ 'the daughter' ACC. is pronounced [tiˈgori] or [tiˈŋgori], then the boundary between /tin/ and /ˈkori/ is labelled *0s*; when the same sequence is pronounced [tin ˈkori] then it is labelled *0m*. In contrast, the sequence /i ˈkori/ 'the daughter' NOM., in which sandhi is not possible, is labelled simply *0*. The *m* diacritic is used with BI 1, 2, and 3, to mark cases in which the transcriber feels that a certain boundary is present, yet the tonal events that normally accompany it are not in place. For

[1] In her review Janet Fletcher suggested that the ip and IP BIs should be 3 and 4 respectively to bring GRToBI in line with MAE ToBI. We decided not to follow her suggestion, because skipping level 2 would make our annotation system less transparent. As it is, GRToBI is similar to other systems, such as Japanese ToBI (Venditti, Ch. 7 this volume, page 172).

example, when the transcriber feels that a sequence which does not end with a phrase accent nevertheless forms a separate ip, then the boundary between this and the following ip should be labelled 2m.

Finally, *p* is used to mark pause at a given boundary, and *?* is used to mark uncertainty about the strength of a boundary. In cases of uncertainty the highest of the two possible candidates is marked, together with a matching analysis in the Tone Tier; if this is not possible (i.e. if the transcriber does not find the tonal cues that normally accompany a particular break index), then *m* should also accompany the break index.

4.5.5. *The Miscellaneous Tier*

The purpose of the Miscellaneous Tier is to encode information about the utterance that is beyond the scope of the other tiers, but may help the users in understanding the information encoded in those utterances. Thus, comments such as disfluency, speaking rate, or pitch-halving (illustrated in Figure 4.9) are marked in this tier.

4.6. DISCUSSION AND CONCLUSION

We have presented here a prosodic analysis of Greek, concentrating mostly on intonation and phrasing, but also dealing (albeit to a lesser extent) with stress and rhythm. It transpires from this analysis that the prosody of Greek is by and large understood, though certain issues remain unresolved. Among them are the phonology and phonetics of downstep, the phonetic realization of accents under tonal crowding, and the phonological representation of the pitch accent currently analysed as H*+L.

Further, what emerges from this analysis is that the prosody of Greek has certain characteristics that merit further consideration. One of these relates to the phonological relevance and modelling of downstep. This issue has been extensively discussed in the literature (for a review, see Ladd 1996), but it is still far from being understood. This is, after all, the reason why downstep is explicitly annotated in many ToBI systems, in contrast to the theoretical works on which these systems are based (cf. Silverman *et al.* 1992; and Beckman and Pierrehumbert 1986, respectively). Yet, certain aspects of downstep are taken for granted, such as that downstep is triggered by bitonal accents. The Greek data clearly show that this is not a universal tendency, and point towards an analysis in which downstep is seen as an independent—rather than predictable—intonational feature (supporting Ladd's 1996 view on this issue).

A related issue is that of scaling influences between phrasal tones: in most intonational systems H- phrase accents upstep L% boundary tones, while L- phrase accents downstep H% boundary tones. Again, Greek does not exhibit this tendency. This means that the description of Greek intonation requires the use of a mid level, namely the tones we analyse as !H- and !H%.[2] This is problematical, given that the autosegmental/metrical framework assumes that intonational patterns can be adequately modelled using only two tones, H and L. However, if we were to limit ourselves to H and L for the description of Greek intonation, we would have to resort to an analysis in which the downstepped patterns involved sequences of abstract L tones, which are not phonetically realized but serve to downstep others (as in Beckman and Pierrehumbert 1986). Apart from the fact that such an analysis would be highly abstract and unmotivated, at present there appears to exist strong evidence against it, since both non-downstepped L- H% and non-upstepped H-L% sequences are attested in Greek.

A third point that emerges from the Greek data is the asymmetric behaviour of phrase accents in final and non-final ip*s*. As noted, phrase accents assume their secondary association only in the former case, and one possible interpretation of this behaviour is that in non-final ip*s* phrase accents have to fulfil their delimitative function, something that is not necessary in final ip*s* (in which boundary tones assume this role). The secondary association of phrase accents is extensively discussed in Grice *et al.* (2000), but it is clear that the puzzle of the phrase accents' behaviour is far from being solved. Data from more languages and possibly from the less common melodies of the languages already studied could provide a better understanding of this issue.

A final point that is worth commenting on is that of phrasing and the prosodic hierarchy. The three-level hierarchy adopted in the GRToBI analysis is the one typically assumed in intonational phonology; at the same time it differs dramatically from the richer hierarchy assumed in prosodic phonology. The reason for the discrepancy may lie in the evidence used: in intonational studies evidence for phrasing comes mainly from tonal patterns, while work in prosodic phonology relies more heavily on sandhi. Greek in this respect may be quite unusual in having a large number of sandhi rules and requiring few levels to account for stress and intonational patterns, thereby bringing to the fore the asymmetries between the two hierarchies. It is still uncertain whether the more elaborate prosodic structures that have been postulated in the past will turn out to be

[2] Carlos Gussenhoven and Esther Grabe (pers. com.) have suggested we use a % marker for mid level pitch, i.e. that we analyse mid level as the absence of a H or L tone. Although this is a suggestion that should be explored further, it does not appear to alter the fact that Greek requires a phonologically contrastive mid level of pitch.

necessary. It this respect, however, it is clear that GRToBI can make a real contribution by providing natural data on sandhi, a phenomenon that is not easily amenable to laboratory testing. This holds also for other aspects of Greek prosody, such as downstep, which will certainly benefit mostly from the examination of prosodically annotated corpora like the GRToBI database.

To conclude, we hope that the prosodic analysis of Greek presented here will serve as the basis for further research, leading to the examination of the problems that emerged during the development of GRToBI and indicated throughout the paper, as well as to the re-evaluation of certain assumptions currently made in the cross-linguistic study of prosody.

APPENDIX I: LABEL ALIGNMENT CONVENTIONS

- The labels for the L+H*, H* and H*+L pitch accents should be aligned with the highest non-spurious Fo point on the accented vowel.
- For the L* accent the lowest Fo point on the accented vowel should be chosen for alignment.
- For the L*+H pitch accent, for which the canonical alignment of both tones is outside the accented syllable, a reliable point early in the accented vowel should be used.
- Phrase accents should be aligned with the relevant BI 2.
- Phrase accent and boundary tone combinations should be aligned with the relevant BI 3.
- The *enclA* and *accdCL* labels should be placed above or below the relevant accent in the Tone Tier.
- The transcriptions in the PrWords Tier should be aligned with the right edge of the sequence of orthographic items (presented in the Words Tier) that form one PrWd.
- Transliterated forms in the Words Tier are aligned at the (acoustic) right edge of the relevant word.
- Break indices are aligned at the (acoustic) right edge of the relevant constituent.

APPENDIX II: PHONETIC TRANSCRIPTION CONVENTIONS

IPA	**ASCII**	**IPA**	**ASCII**	**IPA**	**ASCII**
p	p	θ	th	ɲ / ɲ̩	N / NN
t	t	ð	D	l / l̩	l / ll
k	k	s	s	r	r
c	c	z	z	ʎ / ʎ̩	L / LL

b / ^{m}b	b / mb	ʃ	$	i	i
d / ^{n}d	d / nd	ç	X	e	e
g / $^{\text{ŋ}}$g	g /Ng	ʝ	j	ɐ	a
ɟ / $^{\text{ɲ}}$ɟ	J / NJ	x	x	o	o
β	B	ɣ	G	u	u
f	f	m/m̩	m/mm	y	y
v	v	n/n̩	n/nn		

In addition to the above symbols, the following conventions should be used:

- Noticeably centralized vowels should be transcribed as @.
- Noticeably nasalized vowels should be transcribed with a following tilde; e.g. a~ for [ɐ̃].
- In cases of vowel coalescence, both vowels should be transcribed and joined by+; e.g. u+o for [o̞] resulting from a sequence of /u/ and /o/.
- Whispered vowels should be transcribed in brackets; e.g. (i) for [i̥].
- Vowels that phonologically form separate syllables but are phonetically manifested as a rising diphthong (on the basis, e.g. of tonal alignment evidence), should be transcribed with the second vowel capitalized; stress should be placed before the diphthong.
- Stress should be marked before the consonant(s) of the stressed syllable, following IPA conventions. (At present we are agnostic as to syllabification, so we suggest that transcribers mark maximal onsets, unless tonal alignment or their own intuitions suggests otherwise.)

APPENDIX III: ROMANIZATION CONVENTIONS

GREEK	Romanization	GREEK	Romanization	GREEK	Romanization
α	a	ν	n	αι	ai
β	v	ξ	x	ει	ei
γ	g	ο	o	οι	oi
δ	d	π	p	ου	ou
ε	e	ρ	r	αυ	ay

ζ	z	σ	s	ευ	ey
η	h	τ	t	μπ	mp
θ	o	υ	y	ντ	nt
ι	i	φ	f	γγ / γκ	gg / gk
κ	k	χ	ch	τσ	ts
λ	l	ψ	ps	τζ	tz
μ	m	ω	w	ντζ	ntz

- When the grapheme combinations that usually represent one vowel (e.g. αι) represent two separate vowels, the graphemes are separated by full stops; e.g. *a.i.d'oni* for *αϊδόνι*.
- Spellings with double graphemes are transliterated in the same way; e.g. *θάλασσα* is transliterated as *th'alassa*.
- In words with more than one syllable, stress is marked as an apostrophe before the stressed vowel. Monosyllables bear no stress mark in the Words Tier.
- Initials capitalized in Greek orthography should be transliterated with capital letters as well.

APPENDIX IV: SUMMARY OF GRToBI LABELS

H* *High pitch accent*: often used as the nucleus in broad focus declaratives and realized as an F0 peak which is preceded by (at most) a small rise.

L* *Low pitch accent*: typically realized as a low plateau.

H*+L *Falling pitch accent*: often used as the nucleus in broad focus declaratives and realized as a fall from high pitch.

L+H* *Rising pitch accent*: in this accent the H is preceded by a noticeable dip and aligns roughly in the middle of the accented vowel; often used to signal narrow focus.

L*+H *Rising pitch accent*: in this accent the L tone typically aligns just before the onset of the accented syllable, while the H tone typically aligns with the beginning of the first post-accentual vowel.

> *Early pitch accent diacritic*: marked before L*+H, it indicates that the accent aligns earlier than typically expected (usual in tonal crowding).

< *Late pitch accent diacritic*: marked before L*+H, it indicates that the accent aligns later than typically expected (usual in tonal crowding).

w *Weak (undershot) pitch accent diacritic*: marked before L*+H it indicates that the L tone is higher than typically expected; marked before L* it indicates that the L is realized as a low point rather than a plateau; both realizations are often found in conditions of tonal crowding.

! *Downstep pitch accent diacritic*: marked before the H tone of a pitch accent it indicates downstep, the lower than expected value of the H tone.

H- *High phrase accent*: marked on the right edge of intermediate phrases.

L- *Low phrase accent*: marked on the right edge of intermediate phrases.

!H- *Downstepped high phrase accent*: marked on the right edge of intermediate phrases; it indicates mid-level pitch.

H% *High boundary tone*: marked on the right edge of intonational phrases.

L% *Low boundary tone*: marked on the right edge of intonational phrases.

!H% *Downstepped high boundary tone*: marked on the right edge of intonational phrases; it indicates mid-level pitch.

---

0 *Break index* showing strongest cohesion; typical of boundaries internal to prosodic words (e.g. boundaries between clitics and hosts); often, but not always, accompanied by sandhi (see diacritic *s).*

1 *Break index* marking prosodic word boundaries: marked on the right edge of a prosodic word followed by another prosodic word within the same intermediate phrase.

2 *Break index* marking intermediate phrase boundaries: marked on the right edge of an intermediate phrase followed by another intermediate phrase within the same intonational phrase.

3 *Break index* marking intonational phrase boundaries: marked on the right edge of an intonational phrase.

enclA *Enclitic accent diacritic*: marked on syllables with enclitic accent (the accent closest to the right boundary of a doubly-accented prosodic word).

accdCL *Accented (pro)clitic diacritic*: marked on words that would normally be cliticized when instead they are accented (and therefore independent prosodic words).

s *Sandhi diacritic*: indicates the presence of sandhi at any prosodic boundary.

m *Mismatch diacritic*: when used with BI 0, it indicates that although the context for sandhi exists, sandhi has not taken place; when used with BI 1, 2 or 3, it marks cases in which the labeller feels a certain boundary is present, but the tonal events that normally accompany it are not in place.

p *Pause diacritic*: it indicates the presence of a pause.

? *Uncertainty diacritic*: it indicates that the labeller is unsure of the strength of a given boundary.

REFERENCES

Arvaniti, A. (1991), 'The Phonetics of Modern Greek Rhythm and its Phonological Implications', Ph.D. dissertation (Cambridge University).

—— (1992), 'Secondary Stress: Evidence from Modern Greek', in G. J. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (Cambridge: Cambridge University Press), 398–423.

—— (1994), 'Acoustic Features of Greek Rhythmic Structure', *Journal of Phonetics*, 22: 239–68.

—— (1997), 'Greek "Emphatic Stress": A First Approach', *Greek Linguistics* 95 (*Proceedings of the 2nd International Conference on Greek Linguistics*), Vol. I (Salzburg: Department of Linguistics, University of Salzburg), 13–22.

—— (2000), 'The Acoustics of Stress in Modern Greek', *Journal of Greek Linguistics*, 1: 9–39.

—— (2001), 'The Intonation of Wh-Questions in Greek', *Studies in Greek Linguistics 21* (Thessaloniki), 57–68.

——, and Baltazani, M. (2000), 'Greek ToBI: A System for the Annotation of Greek Speech Corpora', in *Proceedings of Second International Conference on Language Resources and Evaluation* (*LREC2000*), Vol. 2 (Athens), 555–62.

—— and Garding, G. (in press), 'Dialectal Variation in the Rising Accents of American English', in J. Cole and J. Hualde (eds.), *Laboratory Phonology* 9 (Berlin: Mouton de Gruyter).

——, and Ladd, D. R. (1995), 'Tonal Alignment and the Representation of Accentual Targets', in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 4 (Stockholm), 220–3.

——, ——, and Mennen, I. (1998), 'Stability of Tonal Alignment: The Case of Greek Prenuclear Accents', *Journal of Phonetics*, 26: 3–25.

——, ——, —— (1999), 'Scaling and Alignment of Pitch Targets in Modern Greek Wh-Question Intonation', Ms, University of Cyprus and University of Edinburgh.

——, ——, —— (2000), 'What Is a Starred Tone? Evidence from Greek', in M. Broe and J. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (Cambridge: Cambridge University Press), 119–31.

——, Ladd, D. R. and Mennen, I. (2006), 'Phonetic Effects of Focus and "Tonal Crowding" in Intonation: Evidence from Greek Polar Questions', *Speech Communication*, 48: 667–96.

——, Ladd, D. R. and Mennen, I. (in press), 'Tonal Association and Tonal Alignment: Evidence from Greek Polar Questions and Contrastive Statements', *Language and Speech.*

——, and Pelekanou, T. (2002), 'Postlexical Rules and Gestural Overlap in a Greek Spoken Corpus', *Recherches en Linguistique Grecque*, Vol. I (Paris: L'Harmattan), 71–4.

Baltazani, M. (2006*a*), 'Focusing, Prosodic Phrasing, and Hiatus Resolution in Greek', in L. Goldstein, D. Whalen, and C. Best (eds.), *Laboratory Phonology 8* (Berlin: Mouton de Gruyter), 473–494.

—— (2006*b*), 'On /s/-voicing in Greek,' in *Proceedings of the* 7th International Conference on Greek Linguistics. http://www-users.york.ac.uk/~lang32.

Baltazani, M. (to appear), 'Intonation of Polar Questions and the Location of Nuclear Stress in Greek', in C. Gussenhoven and T. Riad (eds.), *Tones and Tunes. Volume II: Phonetic and Behavioural Studies in Word and Sentence Prosody* (Berlin: Mouton de Gruyter).

—— (in press), 'Intonation and pragmatic interpretation of negation in Greek', *Journal of Pragmatics.*

——, and Jun, S.-A. (1999), 'Focus and Topic Intonation in Greek', in *Proceedings of the XIVth International Congress of Phonetic Sciences*, Vol. 2 (San Francisco), 1305–08.

Beckman, M. E., and Ayers-Elam, G. (1997), *Guidelines for ToBI Labeling* (Columbus, OH: Ohio State University Research Foundation).

——, and Pierrehumbert, J. B. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–310.

——, Hirschberg, J., and Shattuck-Hufnagel, S. (this volume Ch. 2), 'The Original ToBI System and the Evolution of the ToBI Framework'.

Botinis, A. (1989), *Stress and Prosodic Structure in Greek* (Lund: Lund University Press).

—— (1998), 'Intonation in Greek', in D. Hirst and A. DiCristo (eds.), *Intonation Systems* (Cambridge: Cambridge University Press), 288–310.

Condoravdi, C. (1990), 'Sandhi Rules of Greek and Prosodic Theory', in S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Interface* (Chicago: University of Chicago Press), 63–84.

Dauer, R. (1983), 'Stress-Timing and Syllable-Timing Reanalysed', *Journal of Phonetics*, 11: 51–62.

—— (1987), 'Phonetic and Phonological Components of Language Rhythm', in *Proceedings of the XIth International Congress of Phonetic Sciences*, Vol. 5 (Tallinn, Estonia, USSR), 447–50.

Drachman, G., and Malikouti-Drachman, A. (1999), 'Greek Word Stress', in H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (Berlin and New York: Mouton de Gruyter), 897–945.

Ellis, L., and Hardcastle, W. J. (1999), 'An Instrumental Study of Alveolar to Velar Assimilation in Fast and Careful Speech', in *Proceedings of the XIVth International Congress of Phonetic Sciences*, Vol. 3 (San Francisco), 2425–8.

Fallon, P. (1994), 'Naturally Occurring Hiatus in Modern Greek', in I. Philippaki-Warburton, K. Nicolaidis, and M. Sifianou (eds.), *Themes in Greek Linguistics* (London: John Benjamins Publishing), 217–24.

Goldsmith, J. A. (1999), 'Introduction' in J. A. Goldsmith (ed.), *Phonological Theory: The Essential Readings* (Oxford: Blackwell Publishers), 1–16.

Grice, M., and Benzmüller, R. (1995), 'Transcription of German Intonation Using Tobi-Tones—The Saarbrücken System', *Phonus*, 1: 33–51.

——, D'Imperio, M., Savino, M., and Avesani, C. (this volume Ch. 13), 'Towards a Strategy for Labelling Varieties of Italian'.

——, Ladd, D. R., and Arvaniti, A. (2000), 'On the Place of Phrase Accents in Intonational Phonology', *Phonology*, 17: 143–85.

——, and Truckenbrodt, H. (2001), 'Hybrid Tones in Optimality Theory', Paper presented at the Holland Institute of Linguistics Phonology Conference 5, (University of Potsdam: January 2001).

Holst, T. and Nolan, F. J. (1995), 'The Influence of Syntactic Structure on [s] to [ʃ] Assimilation', in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (Cambridge: Cambridge University Press), 315–33.

Joseph, B. D., and Philippaki-Warburton, I. (1987), *Modern Greek* (London: Croom Helm).

Kaisse, E. M. (1985), *Connected Speech: The Interaction of Syntax and Phonology* (New York: Academic Press).

Ladd, D. R. (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

Liberman, M., and Pierrehumbert, J. (1984), 'Intonational Invariance Under Changes in Pitch Range and Length', in M. Aronoff and R. Oehrle (eds.), *Language Sound Structure: Studies in Phonology Presented to Moris Halle by his Teacher and Students* (Cambridge, MA: MIT Press), 157–233.

Malikouti-Drachman, A., and Drachman, G. (1981), 'Slogan Chanting and Speech Rhythm in Greek', in W. Dressler, O. Pfeiffer, and J. Rennison (eds.), *Phonologica 1980* (Innsbruck), 283–92.

——, —— (1992), 'Greek Clitics and Lexical Phonology', in W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, and J. R. Rennison (eds.), *Phonologica 1988* (Cambridge: Cambridge University Press), 197–206.

Nespor, M., and Vogel, I. (1986), *Prosodic Phonology* (Dordrecht: Foris).

——, —— (1989), 'On Clashes and Lapses', *Phonology*, 6: 69–116.

Pierrehumbert, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

—— (1981), 'Synthesizing Intonation', *Journal of the Acoustical Society of America*, 70: 985–95.

——, and Beckman, M. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

——, and Hirschberg, J. (1990), 'The Meaning of Intonational Contours in the Interpretation of Discourse', in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271–311.

Revithiadou, A. (1998), *Headmost Accent Wins: Head Dominance and Ideal Prosodic Form in Lexical Accent Systems* (The Hague: Holland Academic Graphics).

Setatos, M. (1974), Fonologia tis Koinis Neoellinikis [*Phonology of Standard Modern Greek*] (Athens: Papazisis).

Silverman, K. E. A., Beckman, M., Pitrelli, J. F., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992), 'TOBI: A Standard for Labeling English Prosody', in *Proceedings of the 1992 International Conference on Spoken Language Processing*, Vol. 2 (Banff, Canada), 867–70.

Venditti, J. J. (this volume Ch. 7), 'The J_ToBI Model of Japanese Intonation'.

Yip, M. (1996), 'Tone in East Asian Languages', in J. A. Goldsmith (ed.), *The Handbook of Phonological Theory* (Oxford: Blackwell Publishers), 476–94.

Zec, D., and Inkelas, S. (1991), 'The Place of Clitics in the Prosodic Hierarchy', in *Proceedings of WCCFL*, 10 (Stanford: SLA), 505–19.

Zsiga, E. C. (1997), 'Features, Gestures, and Igbo Vowel Assimilation: An Approach to the Phonology/Phonetics Mapping', *Language*, 73: 227–74.

5

Transcription of Dutch Intonation

Carlos Gussenhoven

5.1. INTRODUCTION

Transcription of Dutch Intonation (ToDI) is a ToBI-like transcription system developed for standard Dutch. A brief introduction to the prosodic structure of this language, which has no lexical tone, can be given on the basis of (1).

(1) (*a*) Utterance (U)
|
Intonational Phrase (IP)
|
Phonological Phrase (PP)

(*b*)

Initial boundary tones:	%T
Pitch accents:	T*(T)
Final boundary tones:	T%

The IP corresponds to the 'tone group' of the British English tradition, which is closest to the IP as used in ToBI, and is the constituent demarcated by %T and T%. The PP is the domain for clash resolution (to be illustrated below);

I am very grateful to Jacques Terken and Toni Rietveld for their competent and supportive collaboration on the design of the ToDI system and the ToDI course. Toni Rietveld deserves special credit for making everybody aware that a system like ToDI was urgently needed, not in the least because of the data collection for the *Corpus Gesproken Nederlands*, a one-million word spoken corpus of Dutch, was about to begin. The initial discussions about what ToDI should be like were conducted with the Nijmegen prosody group of 1998, and I thank Esther Grabe, Judith Haan, Joop Kerkhoff, Erwin Marsi, Anneke Neijt, Brechtje Post, and Felix Vieregge for their sensible comments and useful contributions. Valuable input was provided by Aditi Lahiri and her colleagues after a presentation of an earlier version at the University of Konstanz in October 1998. Ludmila Menert and Arthur Dirksen of Fluency Speech Technology (Utrecht) are responsible for the implementation of the first edition of ToDI as an interactive course, which was made available on the Internet in July 1999, while Joop Kerkhoff produced the programs for the second edition (2004) < http://todi.let.kun.nl>. Their work was supported by a development grant from the Netherlands Foundation for Scientific Research (NWO).

there is no intermediate constituent between PP and IP. The U is the highest constituent. A minimal instantiation of the elements in (1) would occur on a one-word utterance, like *GefeliciTEERD* 'Congratulations', *HalLO* 'Hello', or JA 'Yes', as exemplified in (2), spoken with a perfunctory intonation.

```
(2)   U{ IP[ PP(JA )PP ]IP }U   'Yes?'
          |     |
         %L     H*
```

Usually, utterances are more complex. In (3), an example is given of a more elaborate expression, with possible lexical instantiations given in (3 *a, b, c*).

```
(3)   U{ IP[ PP( *     * )PP PP( *     *     * )PP ] IP  IP[ PP ( *    * )PP ]IP}U
      |          |     |         |     |     |     |       |      |    |     |
      %L         H*L   H*L       H*L   H*L   H*    H%      %L     H*  !H*L   L%
```

(*a*) De NIEUwe archiTECT / bleef KRAP ZES MAANDEN, maar NIEmand vond dat een proBLEEM
'The new architect stayed less than six months, but nobody minded.'

(*b*) De TWEEde KEER / kwam er een RAAR, KLEIN MANnetje kijken, in een PAARS PAK
'The second time a strange little man came to watch, in a purple suit.'

(*c*) Mijn EIgen DOCHter / staat BOven ELke verDENking, zoals u VAST al had verWACHT
'My own daughter is above all suspicion, as you had no doubt expected.'

As these examples suggest, there are no prosodic restrictions on which words appear in which positions. Except for function words with schwa, which are only accented in metalinguistic usage and a few rare idioms, all words are accentable. There are a number of reasons why words are not accented:

(i) *Morphosyntactic reasons*: right-hand constituents of noun and verb compounds are deaccented, for instance. In addition, certain syntactic constituents that are appended at the right edges of clauses typically remain unaccented. Among these are general place and time adverbials, like *hier* 'here' and *vandaag* 'today', vocatives, and reporting clauses, like *zei Jan* 'said John'. Also, the verb is unaccented when it is adjacent to one of its arguments in an 'eventive' sentence. These conditions may co-occur in the same sentence, as shown in (4). While fully grammatical, this is a somewhat contrived example, as long stretches of unaccented speech are avoided. *Postbode* is a compound, literally 'post-messenger', *gisteren* a general place adverbial, *heeft* *een ongeluk gehad* the predicate of *postbode, Hendrik* a vocative, and *zei Rie op veelbetekenende toon* a reporting clause.

(4) 'De POSTbode heeft gisteren een ongeluk gehad, Hendrik', zei Rie op veelbetekenende toon
the postman has yesterday an accident had Henry said Mary on knowing tone
'The postman had an accident yesterday, Henry', said Mary knowingly

(ii) *Focus structure*: the second reason for deaccentuation is that, within the IP, the focus of the sentence is followed by unaccented words, causing the last pitch accent of the intonational phrase to mark the right edge of the focus constituent. This is why the only accent in (5) is on *niets* 'nothing'.

(5) [A: Welke gordijnen passen beter bij de bank?]
['Which curtains would better match the colour of the couch?']
B: NIETS past er bij de kleur van de bank!
'NOTHING matches the colour of the couch!'

(iii) *Phonological*: like French and English, the phonological phrase governs the distribution of pitch accents through rhythmic considerations. Rhythm-induced deletions and additions of pitch accents are particularly relevant for adjective compounds. For instance, *tweedehands* 'second-hand' has a pitch accent on the second member *hands* when this is the last pitch accent in its phonological phrase, as in (6*a*). The context for this expression must be a conversation about *meubelen* 'furniture' to explain the lack of accent on this word. Some adjective compounds tend to have pitch accents on both constituents in this type of position, in particular when occurring in IP-final position, like *ijskoud* 'ice-cold', *wonderschoon* 'wondrously beautiful', *bloedmooi* 'stunningly beautiful', as shown in (6*b*). When another pitch accented word follows in the same PP, as in (6 *c*, *d*), only the first constituent is accented, a pronunciation known as 'stress shift'. The phenomenon is not restricted to compound words (cf. Visch 1999 and references therein; instead of medial PP-boundaries, there may additionally be IP-boundaries in (6*a*) and (6*c*)).

(6) (*a*) U { IP [PP(TweedeHANDS meubelen) PP PP(kunnen we niet LEveren) PP] IP } U
'SECOND-HAND pieces of furniture we cannot SUPPLY'

(*b*) U { IP [PP(Ze is BLOEDMOOI) PP] IP } U
'She is STUNNINGLY BEAUTIFUL'

(*c*) U { IP [PP (Een TWEEDEhands TAfel)PP PP (stond bij het RAAM) PP] IP } U
'A SECOND-HAND TABLE stood by the WINDOW'

(*d*) U { IP [PP(Een BLOEDmooi MEISje)PP] IP } U
'A STUNNINGLY BEAUTIFUL GIRL'

The prosodic hierarchy of Dutch continues downward with the prosodic word, the foot, and the syllable (Booij 1995: 143). Not all of these are relevant to intonational structure. The foot, a trochee, is relevant because its head serves as the association site for the pitch accents of the language. Outside 'stress shift' contexts (cf. (6)), one particular foot in a phonological word will attract a pitch accent, and its head is the word stress (cf. Hayes 1995: 18ff; Gussenhoven and Bruce 1999). Recent treatments of Dutch word stress are Zonneveld *et al.* (1999: 492–515) and Gussenhoven (2004).

5.1.1. *Melodic aspects*

The melodic aspects of Dutch intonation are quite complex. Although the number of pitch accents in a PP will rarely exceed three, there is no principled limit, as there is in Bengali, which allows only one (Hayes and Lahiri 1991), and thus no limit to the number of pitch accents in the IP. Moreover, the language offers a large number of pitch accents to choose from, also in prenuclear position, which express various discoursal meanings. In this latter respect, it is different from French, which has a choice of only two pitch accents in prenuclear position, or Bengali, which must choose L* H_P (Hayes and Lahiri 1991; Post 2000). There are constraints on the number of *different* pitch accents within the IP: if there are more than two, the prenuclear ones are usually the same.

The primary purpose of ToDI is to make a transcription system available for characterizing the intonation of Dutch utterances or example sentences, including large amounts of spontaneous speech, from which subsequently generalizations about accentuation, phrasing, and tone choice could be extracted. It represents an improvement over the notation proposed in the IPO grammar (Collier and 't Hart 1981; 't Hart *et al.* 1990; 't Hart 1998) in that many more contrasts are catered for. It is more transparent than the description of Gussenhoven (1988, 1991). The same advantage is held over Beckman and Pierrehumbert's (1986) description of American English, from which the ToBI system was derived (Beckman and Ayers 1994). For instance, ToDI uses no abstract tones of the sort L-H% to mean 'mid level pitch', but has only tones for which individual phonetic targets can be identified.

Before presenting the system, a number of features are highlighted in Section 5.2 that distinguish it from ToBI. Section 5.3 then presents and exemplifies the main nuclear and pre-nuclear contours. In Section 5.4, a brief account is given of contours that are predicted by ToDI, if it is viewed as an orthogonal system, but of which we have no recorded examples. Finally, Section 5.5 systematically compares ToDI notation with the notation used in the IPO grammar ('t Hart *et al.* 1990), my own earlier description, and American English ToBI.

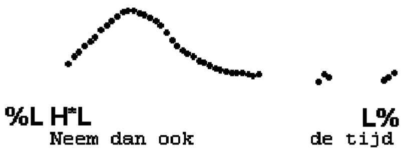
5.2. SOME LESS COMMONLY ADOPTED CONVENTIONS

ToDI has a number of features that may be unexpected for users of other ToBI-like systems. First, the system only covers the 'To' part of ToBI: prosodic breaks are only included from the tonally marked constituent (the IP) onwards. Of course, some version of the ToBI break indices can always be combined with the ToDI transcription. BI3, ToBI's ip-boundary, can be used for the PP-boundary, and BI1 and BI2 can be used as in ToBI. Second, the final boundary tones of the IP, the only tonally marked constituent in the system, are optional. With two tones, this leads to a three-way opposition in the way IPs end, as opposed to a four-way opposition when a two-layered structure with obligatory tones is assumed. Third, some extra-sentential constituents form accentless IPs. Fourth, there are no leading tones in pitch accents. Fifth, the tones of bitonal pitch accents need not be realized close together in time: prenuclear bitonal pitch accents may define quite long gradual slopes, depending on the distance to the next pitch accent. And lastly, singleton pitch accents H* and L* describe high level and low level pitch, respectively, rather than a single high and low target from which to rise or fall. Below, we discuss each point in turn.

5.2.1. *A single tonally marked phrase*

The intonational phrase is the only prosodic constituent to be marked by boundary tones.[1] Its beginning is always marked by a boundary tone, like L% (see Section 5.2.2). At the end, the boundary tones L% and H% may appear. Example (7) illustrates the occurrence of %L and L% with the pitch accent H*L, the most neutral contour of the language (cf. 't Hart *et al.* 1990).

[1] The constituent was referred to as the Association Domain (AD) in my own work (e.g. Gussenhoven 1990), since the intonationally defined constituent does not always coincide with the Intonational Phrase as defined by other criteria. This complication is ignored here.



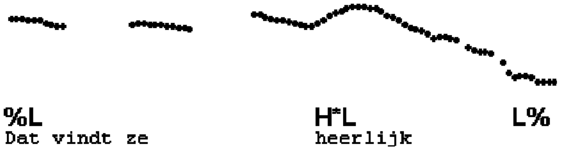
(7) TAKE then so the time
'So then take the time'

In having only a single tonally marked constituent, ToDI differs from the two-phrase intonational structure proposed by Beckman and Pierrehumbert (1986) for English, with an intermediate phrase and an intonational phrase both contributing final boundary tones. In van den Berg *et al.* (1992) and Gussenhoven and Rietveld (1992) it was assumed that, in addition to the IP, the Utterance was provided with boundary tones in Dutch. Specifically, any occurrence of H*L L% and L*H H% was taken to mark the end of an Utterance, while H*L H% was taken to mark the IP-end. The effect was that only Utterance-final IPs could end in extra low pitch (the effect of LL%) or extra high pitch (the effect of HH%). This assumption was made in the interest of a synthesis-by-rule system for intonation, NIROS, which simulates intonation contours for read speech and which necessarily represents an idealization of speaker behaviour. However, in spontaneous speech, there is no indication that the regularity holds. For German, which is closely related to Dutch, Grabe (1998: 171) found HH% as readily in final as in non-final IPs.

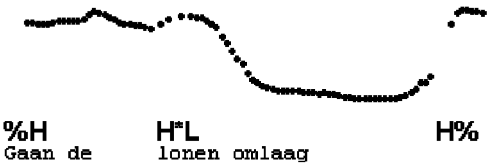
5.2.2. *Boundary tones*

ToDI recognizes three ways in which unaccented syllables can be pronounced at the beginning of the IP. One of these is mid or low pitched, and clearly represents the neutral option; a second is high pitched, a marked option, though it is not uncommon before low pitched accents. The third is a rare, highly marked falling pattern. These patterns are transcribed as in (8), and illustrated in (9), (10) and (11).

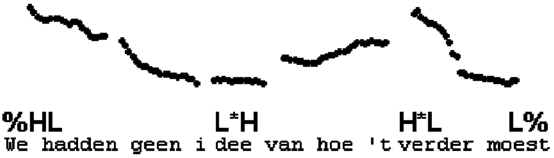
(8) Initial boundary tones: %L
%H
%HL



(9) that finds she LOVELY
'She really likes that!'

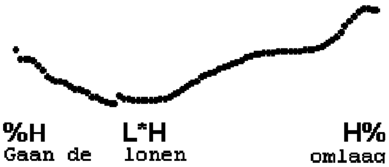


(10) go the WAGES down
'Are wages going down?'



(11) we had no IDEA of how it FURTHER must-PAST
'We had no idea how to go on'

There are two final boundary tones, H%, illustrated in (10) and (12), and L%, illustrated in (7), (8) and (11) above.



(12) go the WAGES up
'Are wages going up?'

Final boundary tones are optional, leading to three phonological right edges of the IP. ToDI's precursor, Gussenhoven (1988), took these boundary tones to be a matter of automatic spell-out, and the nuclear contours in (7) and (12) were simply transcribed H*L and L*H, respectively. It was only when this description was implemented in NIROS (cf. Gussenhoven and Rietveld 1992) that these effects came to be represented by means of separate tones, and the absence of these tones came to be the representation of 'half-completion', which would occur in the mid-ending counterparts of (7) and (11), illustrated in (17) and (24) below. An example of the contrastive absence of a boundary tone from Gussenhoven and Rietveld (1992) is reproduced in Figure 5.1. The advantages of making T% optional were convincingly argued for by Grabe (1998) for German and English, who introduced the notation 0% to signal absence of tone at an IP-boundary.

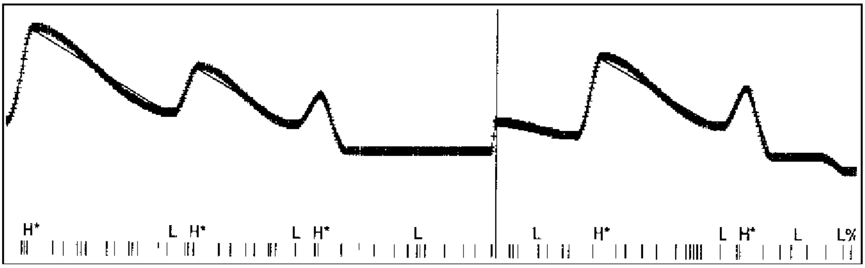
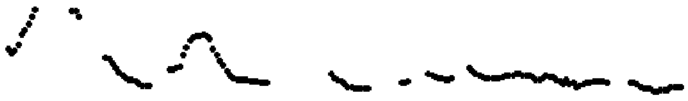


FIGURE 5.1 Example of a rule-based synthesized contour on *PTT Telecom Inlichtingen Telefoonnummers Binnenland* '*PTT* Telecom Information Telephone-numbers Interior', consisting of five H*L pitch accents in two IPs with downstepping accents, one with H*L in Utterance non-final position and one with H*L L% in Utterance-final position. From Gussenhoven and Rietveld (1992).

5.2.3. *Accentless IPs*

Utterance-final IPs may be accentless. Informal observation in a limited corpus suggests that these IPs often express some reformulation of a previous IP, or contain the reporting clause after a direct quotation. This confirms earlier claims that IPs need not contain pitch accents (Trim 1959; Pierrehumbert 1980: 101; Gussenhoven 1990). The pronunciation of such accentless IPs would appear to be determined by the way the preceding IP ends. The tones that occur after the last T* of the preceding IP are repeated in the accentless IP, with a trailing tone (if any) marking the initial boundary, and any other tones marking the final boundary. An example is given in (13).



%L H*L !H* !H*L L% L L%
Netjes met drie woorden spreken zegt m'n moeder altijd

(13) DECENTLY with THREE WORDS speak says my mother always
'"Always speak with three words", my mother always says'
(The expression 'speak with two words' is equivalent to
'say please')

In an attempt to express the clitic-like status of such accentless IPs, their initial boundary is not provided with a %. In that way, we do not compromise the status of the tone transcribed at that boundary, which is after all a copy of a trailing tone, not of a boundary tone.

5.2.4. *Lack of leading tones*

Leading tones in pitch accents might exist in Dutch, though they must be rare if they do. Grice (1995: 197) gives an example (her (23)) for British English, and Kohler (1990) investigates the corresponding German contour, called the 'early peak', whose meaning he identifies as 'established'. If used on Dutch *Met de TREIN* [with the train 'By train'], *met* would have low pitch, *de* high pitch, while a fall from mid to low or low pitch, exactly as for downstepped !H*, would occur on *trein*. The pitch accent would be transcribed as H+!H*.

In other ToBI-like systems, leading tones are used for many more purposes than characterizing the high pitch of a pre-accentual syllable. The main use to which these leading tones are put is taken care of by Tone Linking in ToDI, as explained in the next section. In not having leading tones, ToDI follows my earlier description, which was indebted to the British tradition (e.g. Halliday 1970; O'Connor and Arnold 1973), in which pitch accents (both 'heads' and 'nuclear tones') were parsed as beginning in the accented syllable, and accordingly that description did without leading tones as a matter of course.

5.2.5. *Prefinal bitonal pitch accents describing gradual slopes*

Tone Linking, which causes the trailing tone of a prefinal pitch accent to be pronounced just before the next pitch accent, takes care of the pitch immediately before pitch accents, a view which has proved useful in descriptions of German and English (Gussenhoven 1983; Féry 1993, cf. 'displacement', which term may be preferable to 'linking', as the latter term is also in use in the sense of 'association', Grabe 1998). Figure 5.2 illustrates the concept graphically for three pitch accents: the first column gives nuclear realizations, the second prenuclear ones (after Gussenhoven 1983).

5.2.6. *Singleton T* describing level pitch*

ToDI's singleton H* and L* describe accented high and low pitch targets, as in other systems, which, unlike what is intended in other systems, continue that pitch target until a new tone is transcribed., i.e. H* and L* 'spread'. Falling and rising pitch are transcribed with H*L and L*H, respectively.

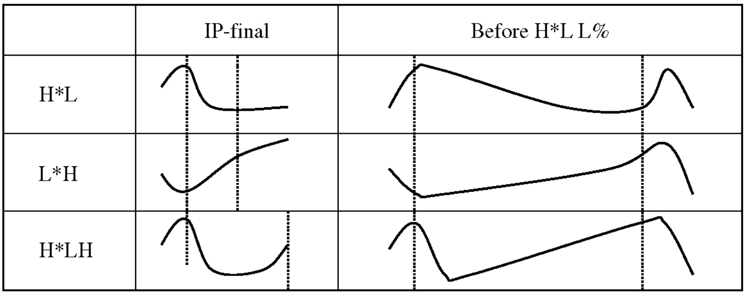


Figure 5.2 Schematic illustration of the effect of (partial) Tone Linking. The dotted lines mark off the contour sections defined by the three pitch accents in final and nonfinal position in the IP (after Gussenhoven 1983).

5.3. ToDI: THE CONTOURS

ToDI was designed to handle all the contrasts presented in 't Hart *et al.* (1990) and Gussenhoven (1988), plus the contrast between 'high rise' and 'low rise' (Gussenhoven and Rietveld 2000). Notation devices for additional features, such as the difference between a normal and a raised peak, can of course be added to the system at the discretion of the researcher. An overview is given in (14).

(14) Initial boundary tones: %L
(repeated from (8)) %H
%HL

Final boundary tones: L%
(optional) H%

Pitch accents: H*
L*
H*L
L*H
H*!H

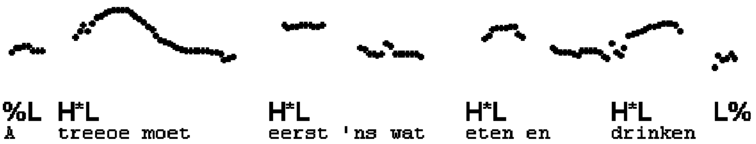
The pitch accents define a sustained high pitch (H*), sustained low pitch (L*), falling pitch (H*L), rising pitch (L*H), and the vocative chant (H*!H). There are modified versions of these pitch accents, given in (15). First, H* and H*L may be downstepped, notated !H* and !H*L. Second, (!)H*L may be prefixed with L*, leading to the tritonal pitch accents L*HL and L*!HL, to be used for (downstepped or non-downstepped) delayed falls and fall-rises (cf. Ladd 1983; Gussenhoven 1983). Lastly, a prefinal steep fall is transcribed H*+L to distinguish it from the gradual prefinal fall. In the second edition, this H*+L has been replaced with H*LH, in line with the analysis in Figure 5.2.

(15) Modified pitch accents !H*
!H*L
L*HL
L*!HL
H*+L (= H*LH in 2nd edition)

5.3.1. *Nuclear contours*

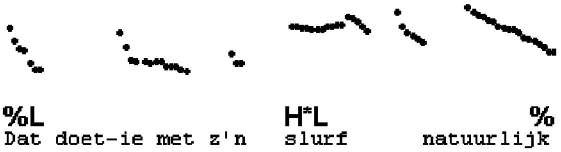
In this section, the most common nuclear contours are described. Somewhat in the manner of O'Connor and Arnold (1973), names will be provided for each contour section defined from the last pitch accent ('nuclear contours').

(i) *The fall*: the fall was already illustrated in (7), (8) and (11). In (16), we give an IP with four falls. The prefinal falls tend to be gradual, the nuclear one steep. The slope of the prefinal fall is not contrastive, and may be steeper without making it sound like a different contour (Collier and 't Hart 1981; cf. also Ladd 1996: 96).



(16) ATREOE must FIRST PARTICLE something EAT and DRINK
'Atreoe had better first eat and drink something'

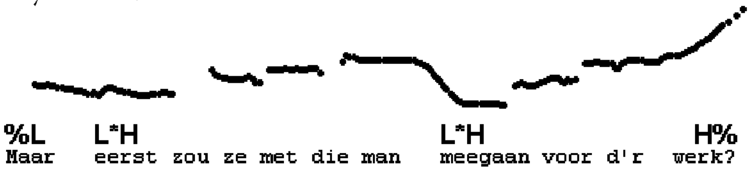
(ii) *The half-completed fall*: when the fall reaches only mid pitch at the IP-end, sounding as if the speaker is less insistent, the fall is half-completed. It is transcribed H*L %, i.e. there is no boundary tone. An illustration is given in (17).



(17) that does-it with his ELEPHANT'S-TRUNK of-course
'It does that with its trunk, of course'

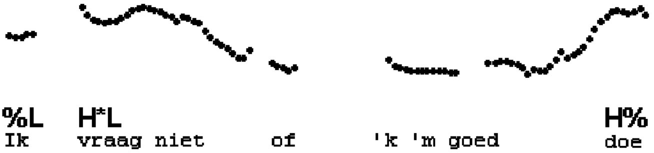
(iii) *The low rise*: example (18) has two L*H pitch accents, followed by H%. Again, the H-tone of the prefinal pitch accent is realized just before the

L* of the next pitch accent, in this case causing a gradual rise, whose slope will vary without obvious implications for the identity of the contour. The part described by L*H H% is termed the 'low rise'.

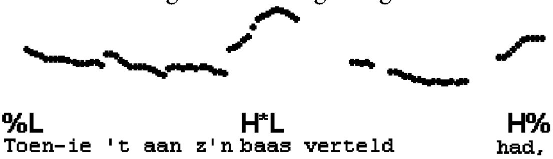


(18) but FIRST would she with that man ACCOMPANY for her work
'But was she first planning to join that man on a business trip?'

(iv) *The fall-rise*: the fall-rise, H*L followed by H%, is a quite frequent contour of Dutch, both in Utterance-final and nonfinal position, as illustrated in (19) and (20), respectively.

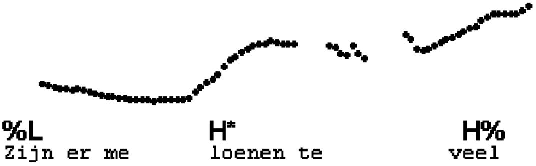


(19) i ASK not whether i it good do
'I'm not asking if I'm doing it right'

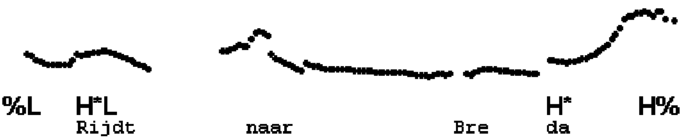


(20) when-he it to his BOSS told had
'When he had told his boss, . . . '

(v) *The high rise*: nuclear rises that begin at mid pitch in the accented syllable are transcribed H* H%. In nonfinal syllables, the pitch is usually low at the beginning of the vowel, and there is a rising movement in the first half of the syllable towards the target of H*. There is high pitch in the IP-final syllable, the target of H%. Two of these 'high rises' are shown in (21), in nonfinal position, and (22), in an IP-final syllable.

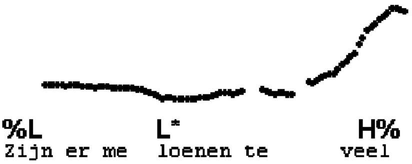


(21) are there MELONS too many
'Are there too many melons?'



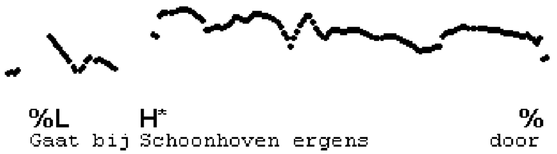
(22) DRIVES to BREDA
'. . . drives to Breda, . . . '

(vi) *The low low rise*: instead of rising immediately, the low target of L* may continue until the IP-final syllable, where the pitch rises. This 'low low rise' is transcribed L* H%, and illustrated in (23).



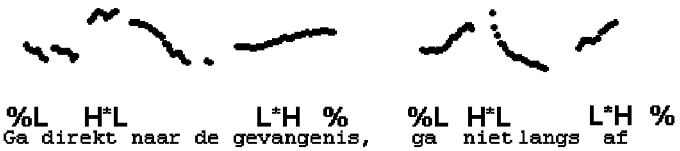
(23) are there MELONS too many
'Are there too many melons?'

(vii) *The level contour*: the most prototypical 'listing' intonation is one that has mid pitch in the accented syllable, which continues at more or less the same level until the IP-end. It is transcribed H* %, and shown in (24).



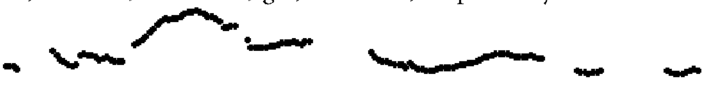
(24) goes near Schoonhoven somewhere through
'. . ., takes a shortcut near Schoonhoven or thereabouts, . . .'

(viii) *The half-completed rise*: like the low rise, the half-completed rise has low pitch in the accented syllable, then rises immediately, but unlike the low rise does not rise further in the final syllable. It is transcribed L*H %, as in (25), where it occurs twice in nonfinal IPs. It is readily usable as a 'listing intonation', as an alternative to the level contour. Phonologically, the difference between L*H H% and H* H% parallels that between L*H % and H* %.



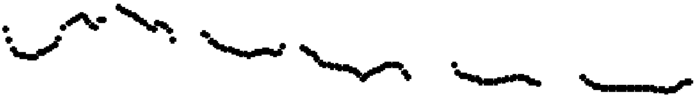
(25) go DIRECTLY to the JAIL go NOT along OFF
'Go directly to jail, don't pass *Go*, . . . '

(ix) *Vocative chant*: the vocative chant is a contour with minimally two pitch levels, frequently accompanied by lengthening of the initial syllables of each level (Gussenhoven 1993). It is transcribed H*!H %, as in (26), where it is used in a nonfinal IP. In multi-level realizations, a contour-type that does not exist in English and forms a cascading sequence of non-accent-lending pitch levels, we simply transcribe H*H!, as in (27), where five levels are formed, on *moet niet me-*, *teen het*, *antwoord*, *ge-*, and *-ven*, respectively.



%L H* !H % %L H* !H*L L%
Als u de bon invult, sturen wij u de gids toe

(26) if you the COUPON fill-in, send WE you the CATALOGUE towards
'If you fill in the coupon, we will send you the catalogue'



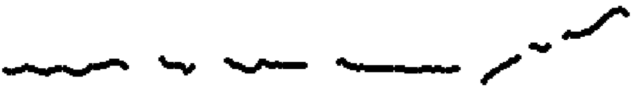
%L H* !H %
Je moet niet meteen het antwoord ge- ven

(27) you MUST not immediately the answer give
'You mustn't answer so quickly!'

5.3.2. *Prefinal pitch accents*

In the first section, we discuss prefinal H*, H*L, L*, and L*H. A separate section is devoted to H*+L.

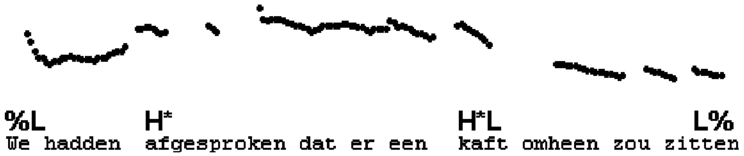
(i) *The prefinal high, fall, low, and rise*: examples with the prefinal fall (H*L) were already presented in (12), (16) and (22), and of the prefinal rise (L*H) in (11) and (18). An example with a prefinal low (L*) is (28). The L* marks the accented syllable as having low pitch, which then continues until the next pitch accent. Understandably, the difference between an unaccented low-pitched syllable and a prefinal accented syllable with L* may be less than obvious. In (28), though, there is a clear sensation of an accent on *werkelijk*.



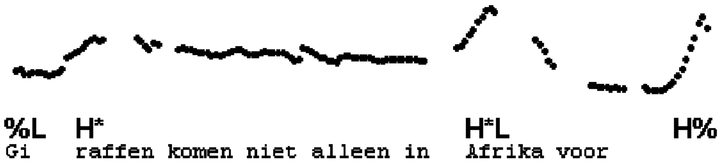
%L L* L*H H%
Zou ze werkelijk zo gek zijn om daarop in te gaan?

(28) would she REALLY be so crazy to there-up IN to go
'Would she really be stupid to accept that invitation?'

Just as L* creates a low-pitched stretch, so H* creates a high-pitched stretch. It commonly appears before H*L, as in (29) and (30). Notice that in (30) the second H* is raised relative to the preceding H*; this feature is not transcribed in ToDI, as it is not clear that contours with and without such raising are phonologically distinct. If required, a symbol like ^ could be used.

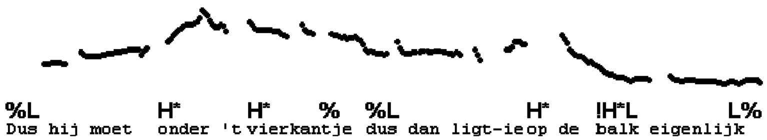


(29) we had AGREED that there a COVER around would be
'We had agreed that you would cover your books'

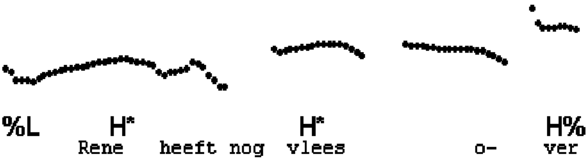


(30) GIRAFFES occur not only in AFRICA VERBAL PARTICLE
'Giraffes not only occur in Africa!'

In the first IP in (31), H* appears before the level tone, H* %. In such contours, the second H* will typically be just a little lower than the first. However, when H* precedes the high rise, H* H%, the second H* is typically just little higher than the first, as illustrated in (32).

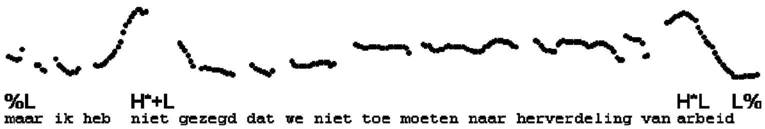


(31) so it must UNDER the SQUARE+DIM so then lies-it ON the BEAM in-fact
'So it goes under the little square, so that it ends up on top of the beam'

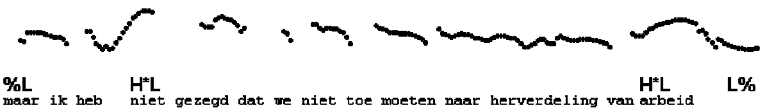


(32) RENÉ has still MEAT left
'René still has meat left'

(ii) *Prefinal fall-rise*: in other systems, the '+' diacritic is standardly used to indicate that tones are grouped together in a bitonal pitch accent. In ToDI, it is used to indicate that the two tones it links are pronounced close together: H*+L thus describes a steep fall. It preempts the convention that the last tone of a polytonal pitch accent moves off to the right, where it creates a target just before the next pitch accent (see Section 5.2.5). The pitch accent H*+L only appears contrastively in prenuclear position, where it is typically followed by a gradual rise to H*. For example, in (33), we have H*+L on *niet* 'not', followed by high pitch on *arbeid*. This contour is clearly distinct from one in which the high pitch on *niet* is followed by a gradual fall, as illustrated in (34), spoken by the author.



(33) but I have NOT said that we not towards must to redistribution of LABOUR
'But I haven't said we mustn't move in the direction of redistribution of labour'

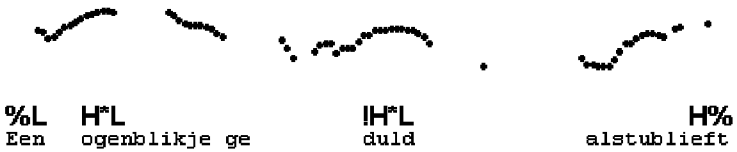


(34)

The explanation of the restrictive distribution of H*+L is that this pitch accent is really the occurrence of the H*LH pitch accent (see Gussenhoven 1983, 1988, and Figure 5.2). This pitch accent resolves as H*L H%, the fall-rise, in IP-final position, but in prenuclear position it is subject to Tone Linking, i.e. undergoes a rightward shift of the last tone, H, as expected. The L-tone, which is nonfinal within the pitch accent, does not shift, and defines the low point of the steep fall from H*. This not only explains why H*+L is followed by a gradual rise, but also why contours like H*+L H*L are readily replaceable with H*L H% %L H*L (cf. Cruttenden 1994: 59), from which they must be historically derived. In (34), the internal IP-break would occur after *gezegd*, the last word of the main clause, which is not the location of the target of L, suggesting L is not some kind of boundary tone. So H*+L is just a more surface-true notation of prefinal H*LH. (In the second edition, the symbol H*LH has been substituted for H*+L.)

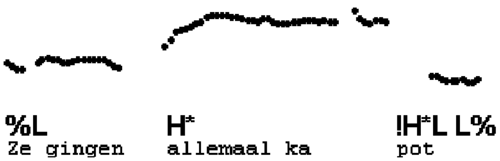
5.3.3. *Downstepped contours*

In (26) and (27), there occurred an example of a downstepped tone, the unaccented trailing tone of the vocative chant H*!H %. As shown in (27), each of the levels is lower in pitch than the one before. The more usual tone to be downstepped is H*. Two types of contour are particularly frequent. In the first, the downstepped H* occurs after H*L, as in (35): the pitch sags between the earlier high peak and the later low peak, which is transcribed by the trailing L-tone of the first pitch accent.



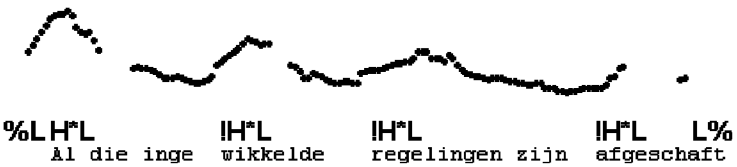
(35) a MOMENT-DIM PATIENCE please
'Just a moment, please'

The second type is probably more frequent, and was termed the 'flat hat' by the IPO intonologists ('t Hart *et al.* 1990). The first accent in (36) has H*, which maintains its pitch until the downstepped !H* of the second accent, causing an early fall which starts just before the final accented syllable. The pitch will variably fall from mid to low or may be low throughout, without there being a noticeable difference. One of the most frequent transcription errors we have encountered is a failure to mark final accents in contours like (35) as accented. In such cases, transcribers incorrectly interpret the contour as one with a single (contrastive) accent on the word which has the first accent.

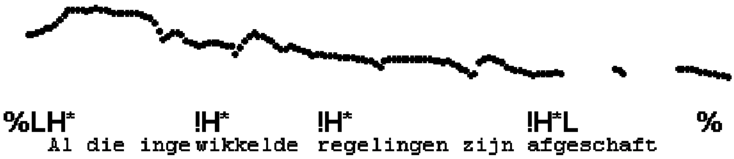


(36) they went ALL BROKEN
'They were all broken'

Examples (37) and (38) show these contours with multiple accents, with cascading series of pitch targets for the downstepping !H*-tones. Example (38) replicates an example in Collier and 't Hart (1981).

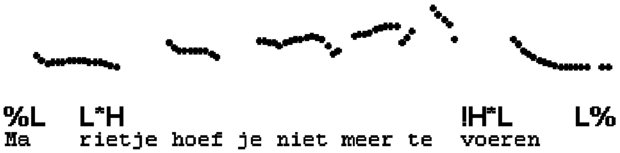


(37) ALL those COMPLICATED RULES have-been ABOLISHED
'All those complicated rules have been abolished'



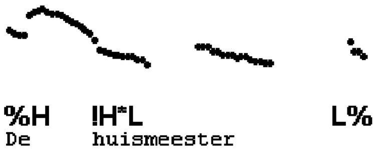
(38)

Downstepped !H* may occur after any pitch accent containing a H-tone. In (39), there is one after L*H.



(39) MARY-DIM need you not anymore to FEED
'Mary you don't have to feed anymore'

Downstepped !H*, lastly, may occur after %H. That is, there are one-accent contours with downstep. This type of contour is not improbable in utterances that represent titles of stories read aloud. Example (40) might be such an utterance.

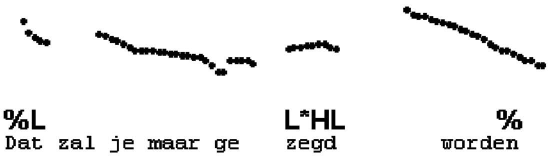


(40) the CARETAKER
'The caretaker'

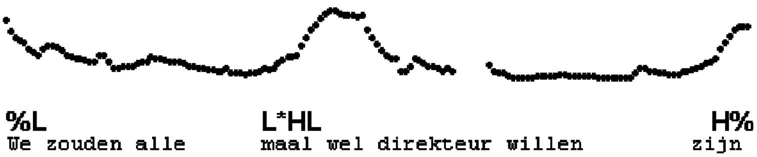
5.3.4. *Delayed accents*

In Gussenhoven (1983), I proposed a modification [DELAY] which can affect the three basic pitch accents of that description, H*L, L*H and H*LH. The delayed version of L*H is a contour that ToDI represents as L* H%, the low rise (cf. Section 5.3.1).[2] The delayed H*L, which creates a low-pitched accented syllable followed by a rise-fall, either inside that syllable if it is IP-final, or in the next syllable if it is not, is transcribed L*HL in ToDI. The extraneous status of the prefixed L* is nicely illustrated by the fact that the original starred tone H* does not get copied into the accentless IPs discussed in Section 2.3. That is, after L*HL H%, for instance, the reported clause has L-H%, not HL-H%.

Examples (41) and (42) illustrate this pitch accent, in combination with different boundary conditions.

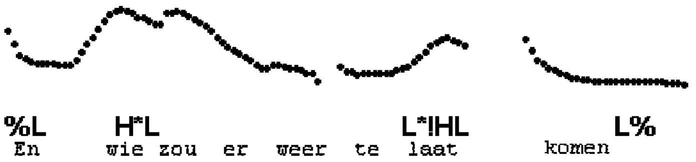


(41) that will you PARTICLE SAID be
'Just imagine that sort of thing being said about you'



(42) we would ALL PARTICLE director want-to be
'Who would all like to be a director'

Of course, delayed falls can also be downstepped. Example (43) shows a downstepped delayed fall-rise.



(43) and WHO would there again too LATE be
'And who should be late again?'

[2] The classification of the delayed peaks and the low low rise in a single morphological category having the meaning of 'significance' was rejected for English by Cruttenden (1994: 117), and has found no support elsewhere. The ToDI transcription with prefixed L* thus follows Ladd's (1980, 1983) classification of 'scooped' or [+delayed peak], which only groups 'rise-fall' and 'rise-fall-rise' of the British tradition together as versions of the 'fall' and 'fall-rise', respectively. As far as I am concerned, the jury is still out on whether the low low rise shares a semantic category with the delayed peaks.

5.4. ToDI'S ORTHOGONALITY

Like Beckman and Pierrehumbert (1986), but unlike American English ToBI, ToDI is an orthogonal system. The number of nuclear contour types is 24, as shown in (44): in the first column, the three parenthesized elements bring the total number of pitch accents to 8.

(44) $$\left\{\begin{array}{c} \mathrm{H}^* \\ (!)\mathrm{H}^*\mathrm{L} \\ \mathrm{L}^*(\mathrm{H}) \\ \mathrm{L}^*(!)\mathrm{HL} \\ \mathrm{H}^*!\mathrm{H} \end{array}\right\}\left\{\begin{array}{r} \mathrm{L}\ \% \\ \mathrm{H}\% \\ \% \end{array}\right\} = 24 \text{ nuclear contours}$$

Above, not all combinations with final boundary conditions in (24) were illustrated, but I believe that all combinations are in fact well formed. The ones that have not been discussed either seem rare, or might have been seen as variants of other contours. Table 5.1 lists the complete inventory that is generated, with labels, in italics if the contour has not been mentioned above.

First, the (delayed or non-delayed) high plateau-slump (the term is from Cruttenden 1994) is a contour that starts high in—or after, in the delayed version—the accented syllable and continues high until a fall to low just before the final syllable. Second, the (falling) low level contours would seem to be used when repeating someone else's statement in a scathing manner. Third, the low-ending vocative chant is a version of the vocative chant that

Table 5.1 Full set of nuclear pitch accents in three boundary conditions generated in the ToDI system, plus prose labels; no examples are available for the italic labels, but it is believed that these represent well-formed contours of Dutch

	L%	H%	%
H*L	Fall	Fall-rise	Half-completed fall
!H*L	Downstepped fall	Downstepped fall-rise	Downstepped half-completed fall
H*	*High plateau-Slump*	High rise	Level
H*!H	*Low-ending vocative chant*	*Vocative fall-rise*	Vocative chant
L*H	*Delayed high plateau-slump*	Low rise	Half-completed rise
L*	*Falling low level*	Low low rise	*Low level*
L*HL	Delayed fall	Delayed fall-rise	Delayed half-completed fall
L*!HL	Downstepped delayed fall	Downstepped delayed fall-rise	Downstepped delayed half-completed fall

expresses a high degree of impatience, while the vocative fall-rise is just like the vocative chant, with a final rise on the final (lengthened) syllable (the latter was given for English in Gussenhoven 1983, as example (36)).

5.5. COMPARISON WITH OTHER SYSTEMS

It would be appropriate at this point to discuss the way other descriptions have dealt with the contours presented in the above sections. Such a discussion not only defines the increase in descriptive coverage, but may also be helpful for those who would like to re-transcribe into the ToDI system contours in the literature, and facilitate the identification of ToDI contours by those who are more familiar with other conventions. However, the space a full treatment would require is prohibitive in the context of this chapter. It is possible, though, to give these comparisons in tabular form, and dispense with a discussion of the advantages and disadvantages of the various systems. This is done in this section.

The first comparison is with the IPO grammar, the second with the autosegmental description (Gussenhoven 1988, 1991; van den Berg *et al.* 1992; Gussenhoven and Rietveld 1992), and the third with ToBI, as proposed for American English (Beckman and Ayers 1994), a language whose intonational system is very similar to that of Dutch.

5.5.1. *The IPO grammar compared with ToDI*

Table 5.2 lists ToDI transcriptions for nuclear pitch accents plus boundary tones that were illustrated in the sections above, together with the IPO transcriptions used to transcribe those contours.[3] It is clear that the IPO grammar has fewer contours, and many ToDI contrasts are simply not expressed (Gussenhoven 1988). As stressed in the latter publication, however, the IPO grammar was never designed to be an exhaustive description. It was based on a corpus of speech, and there is of course no guarantee that in any given corpus all possible contours are actually attested. The IPO researchers also put a lower limit of 6 per cent on the frequency of occurrence of any one contour for it to be included in their grammar.

In Table 5.2, some gaps in the IPO column have been filled with transcriptions that 't Hart, Collier and Cohen (1990) do not give, but that could be used to represent the contours concerned. These non-standard transcriptions

[3] These tables were drawn up in consultation with Jacques Terken. I remain responsible for the interpretations given in them.

Table 5.2 Nuclear contours in ToDI compared with their counterparts in IPO, with prose labels as used in this chapter; bracketed IPO transcriptions are not produced by the IPO grammar, but might be if it were adapted

ToDI		IPO	Prose label
H*L	L%	A	Fall[a]
H*L	%	(1 C)	Half-completed fall
H*L	H%	A 2	Fall-rise
!H*L	L%	A	Downstepped fall[a]
!H*L	%	—	Downstepped half-completed fall
!H*L	H%	—	Downstepped rise-fall
H*	%	1	Level tone
H*!H	%	1 E	Vocative chant
H*	H%	1 2	High rise[b]
L*H	H%	(3 2)	Low rise
L*H	%	(3)	Half-completed rise
L*	H%	(2)	Low low rise[c]
L*HL	L%	(3&B)	Delayed fall
L*HL	%	3 C	Half-completed delayed fall
L*HL	H%	(3&B 2)	Delayed rise-fall
L*!HL	L%	—	Delayed downstepped fall
L*!HL	%	—	Delayed downstepped half-completed fall
L*!HL	H%	—	Delayed downstepped fall-rise[d]

[a] IPO ‘A’ corresponds with H*L L% in all contexts except in the ‘flat hat’ contour, which is transcribed ‘1 A’. This contour is equivalent to H* !H*L L%, i.e. with downstepped H*, and hence phonetically with an earlier fall. In ’t Hart *et al.* (1990), ‘B’ is often used for the late variant of ‘A’. The ‘pointed hat’ is ‘1&A’, and is equivalent to ToDI %L H*L L%.

[b] IPO ‘1 2’ is H* H% rather than L*H H%, both in the description and in the examples given in Collier and ’t Hart (1981), while IPO ‘1’ is L*H % in Collier and ‘t Hart (1981). H* % and L*H % could be distinguished in IPO as ‘1’ and ‘3’, if ‘3’ were used for H* %.

[c] IPO ‘2’ is described as a non-accent lending final rise. The grammar in ‘t Hart *et al.* (1990) generates a singleton ‘2’, however, which could be used to describe this contour.

[d] None of the contours with italic labels in Table 5.1 could readily be accommodated in IPO notation. They were omitted from Table 5.2.

are given in parentheses. The notations for the prenuclear contours are compared in a similar fashion in Table 5.3.

5.5.2. *The earlier autosegmental description of Dutch compared with ToDI*

ToDI bears a strong resemblance to my earlier description. The only contrast recognized in that description which cannot be reproduced in ToDI is that

Table 5.3 Pre-nuclear contours in ToDI compared with their counterparts in IPO; Bracketed IPO transcriptions are not produced by the IPO grammar, but might be

ToDI	IPO	Prose label
H*	1	Prenuclear high
!H*	E	Downstepped prenuclear high
H*L	1D	Prenuclear fall
!H*L	—	Downstepped prenuclear fall
L*	0[a]	Prenuclear low
L*H	4	Prenuclear rise
L*!HL	—	Downstepped delayed prenuclear fall
H*+L	1&A 4 5[b]	Prenuclear fall rise

[a] IPO cannot mark low-pitched accents in low-pitched surroundings, since it is based on movements.

[b] The '5' represents a (non-obligatory) small extra rise above the usual level just before the next accent, which could be interpreted as the movement from the target of H to that of H*, if H*+L is replaced with H*LH of Gussenhoven (1988). That is, IPO '1&A 4' would equally be transcribed H*+L. As Jacques Terken (personal communication) points out, IPO could also use '5' to describe the heightened peak in (30), a feature for which ToDI has no transcription.

between a plain fall-rise and a half-completed fall-rise. The half-completed fall-rise has mid pitch, instead of low pitch, between the high pitch of H* and the high pitch of H%, something that gives the utterance a tentative ring. This contrast was available in Pierrehumbert (1980) as H* L-H% (plain) vs. H*+L H-H% (half-completed), but no provision for it exists in ToBI. It is indeed arguable whether it should not be explained as phonetic variation.

The speech synthesis programme NIROS was designed to take tones and modifications as input in a rather abstract form, and produce a surface phonological representation after the application of automatic spell-out rules (Gussenhoven and Rietveld 1992). As shown in Table 5.4, ToDI is very close to these latter representations. The modifications were features on the IP, and had an effect on some or all of the pitch accents in it. Thus, > delays all accents, = produces a half-completed version of the last pitch accent, ! causes all H*'s except the first to be downstepped, and &, 'narration', causes all T*'s to spread right, pushing the trailing tone up against the next T* or the IP-boundary. (No provision was made for 'stylization', which requires additional durational manipulation.) For instance, input representation (45) would be transformed into (46), which latter representation feeds the phonetic implementation. The contour has a gradual slope from the first accent (pre-nuclear rise), followed by an early fall on the second (downstepped fall).

TABLE 5.4 Nuclear contours in ToDI compared with their counterparts in the earlier autosegmental description with prose labels as used in this chapter

ToDI	Autosegmental	Spell out	
H*L L%	(...H*L)	H*L L%	Fall
H*L %	=(...H*L)	H*L %	Half-completed fall
H*L H%	(...H*L)	H*L H%	Fall-rise
!H*L L%	! (...H*L)	!H*L L%	Downstepped fall
!H*L %	=! (...H*L)	!H*L %	Downstepped half-completed fall
!H*L H%	! (...H*LH)	!H*L H%	Downstepped rise-fall
H* %	stylized L*H	*no impl.*	Level tone
H*!H %	stylized H*L	*no impl.*	Vocative chant
H* H%	—	—	High rise[a]
L*H H%	L*H	L*H H%	Low rise
L*H %	=(...L*H)	L*H %	Half-completed rise
L* H%	&(...L*H)	L* H %	Low low rise[b]
L*HL L%	>(...H*L)	L*HL L%	Delayed fall
L*HL %	>=(...HL*)	L*HL %	Half-completed delayed fall
L*HL H%	>(...H*LH)	L*HL H%	Delayed rise-fall
L*!HL L%	>!(...H*L)	L*!HL L%	Delayed downstepped fall
L*!HL %	>=!(...HL*)	L*!HL %	Delayed downstepped half-completed fall
L*!HL H%	>!(...H*LH)	L*!HL H%	Delayed downstepped fall-rise

[a] The high rise was incorrectly considered a variant of the low rise (cf. Gussenhoven and Rietveld 1997).

[b] The delayed rise can undergo more or less right-shifting of the rising movement (Gussenhoven 1983). In its most extreme form, it is equivalent to the 'narrated' L*H, which was described as resulting from spreading L* to the IP-end, where it left some space for H. Delay was implemented in NIROS by inserting an extra L-tone before the L* (or H*, in the case of H*L) causing a right-shifting of some 100 ms.

(45) {!(L*H H*L)}

(46) {%L L*H !H*L L%}

The main addition to the earlier description is H* H%, i.e. the contrast between the low rise and the high rise. Among the pre-nuclear pitch accents, H*+L corresponds to H*LH, as explained in Section 4.2.2. Of the contours with italic labels in Table 5.1, the earlier description has no way of creating the low-ending vocative chant. Neither could it produce low level contours. It described the high-plateau slump by means of the narrated H*L, and the vocative chant as the stylized fall-rise. Neither of these was implemented.

TABLE 5.5 Nuclear contours in ToDI compared with their counterparts in ToBI

ToDI		ToBI		Prose label
H*L	L%	H*	L-L%	Fall
H*L	%	H*	H-L%	Half-completed fall
H*L	H%	H*	L-H%	Fall-rise
!H*L	L%	!H*	L-L%	Downstepped fall[a]
		H+!H*	L-L%	
!H*L	%	!H*	H-L%	Downstepped half-completed fall
!H*L	H%	!H*	L-H%	Downstepped rise-fall
H*	%	H*	H-L%	Level tone
H*!H	%	H*	!H-L%	Vocative chant
H*	H%	H*	H-H%	High rise
L*H	H%	L*	H-H*	Low rise
L*H	%	L*	H-L%	Half-completed low rise
L*	H%	L*	L-H%	Low low rise
L*HL	L%	L*+H	L-L%	Delayed fall
L*HL	%	L*+H	H-L%	Half-completed delayed fall
L*HL	H%	L*+H	L-H%	Delayed rise-fall
L*!HL	L%	L*+!H	L-L%	Delayed downstepped fall
L*!HL	%	L*+!H	H-L%	Delayed downstepped half-completed fall
L*!HL	H%	L*+!H	L-H%	Delayed downstepped fall-rise

[a] ToBI has two transcriptions for the downstepped fall. The leading H is used to describe a high plateau that runs from a preceding H-tone to the syllable before the downstepped accent. This implies that a contour with downstepping level pitches is distinct from one in which the pitch slithers down from one (!)H* to the next, following the general ToBI convention of interpolating between one tonal target and the next. ToDI does not make this assumption.

5.5.3. *American English ToBI compared with ToDI*

The intonation systems of English and Dutch are extremely similar. American English ToBI might therefore have been used for the transcription of Dutch, as suggested by the correspondences in Table 5.5. There are two reasons for nevertheless proposing a new system. One is that Dutch has contours which American English lacks (and for which thus no transcription is available in ToBI), like the high-plateau slumps (Table 5.1), or for which ToBI has proposed no transcription, like the low-ending vocative chant. The second is that employing two phrase-types with obligatory right-hand boundary tones has a number of drawbacks, such as creating undesirable intermediate phrase breaks (cf. Ladd 1996: 96 ff.; Nolan and Grabe 1997), which ToDI avoids.

5.6. CONCLUSION

ToDI is the most complete transcription system ever to be proposed for Dutch. It was developed in collaboration with Toni Rietveld and Jacques Terken, as well as a number of researchers in the area of prosody who took part in the Nijmegen prosody group in 1998. It is largely based on my autosegmental description of Dutch (Gussenhoven 1988, 1991; van den Berg *et al.* 1992), more particularly on the 'spelled-out' version as used in the Nijmegen synthesis-by-rule programme NIROS (Gussenhoven and Rietveld 1992).

In order to make the system accessible and learnable, an interactive course was developed for the Internet in collaboration with Jacques Terken and Toni Rietveld, and designed by Ludmila Menert and Arthur Dirksen of Fluency Speech Technology in Utrecht. After a brief trial period by users of a CD-Rom containing a first version, the course was made freely accessible on the Internet in July 1999. A second edition, which includes a synthesis facility, was posted in 2004 (http://todi.let.kun.nl). The synthesis facility allows users to compare the intonation contour of the original speech file with that of a synthesized version, and on the basis of this comparison, adjust their initial transcription.

REFERENCES

Beckman, M. E., and Ayers, G. M. (1994), *Guidelines for ToBI Transcription* (version 2.0, February 1994).

——, and Pierrehumbert, J. B. (1986), 'Intonational Structure in English and Japanese', *Phonology Yearbook*, 3: 255–310.

Berg, R. van den, Gussenhoven, C., and Rietveld, T. (1992), 'Downstep in Dutch: Implications for a Model', in G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (Cambridge: Cambridge University Press).

Booij, G. E. (1995), *The Phonology of Dutch* (Oxford: Oxford University Press).

Collier, R., and 't Hart, J. (1981), *Cursus Nederlandse Intonatie* (Leuven: Acco).

Cruttenden, A. (1994), *Intonation*, 2nd edn. (Cambridge: Cambridge University Press).

Féry, C. (1993), *German Intonational Patterns* (Tübingen: Niemeyer).

Grabe, E. (1998), *Comparative Intonational Phonology: English and German*, Ph.D. dissertation (University of Nijmegen) (Nijmegen: MPI Series in Psycholinguistics 7).

Grice, M. (1995), 'Leading Tones and Downstep in English', *Phonology*, 12: 183–233.

Gussenhoven, C. (1983), 'A Semantic Analysis of the Nuclear Tones of English' (Indiana: IULC) included as Chapter 6 in C. Gussenhoven (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

GUSSENHOVEN, C. (1988), 'Adequacy in Intonation Analysis: The Case of Dutch', in H. van der Hulst and N. Smith (eds.), *Advances in Nonlinear Phonology* (Dordrecht: Foris), 95–121.

—— (1990), 'Tonal Association Domains and the Prosodic Hierarchy in English', in S. Ramsaran (ed.) *Studies in the Pronunciation of English* (London: Routledge), 27–37.

—— (1991), 'Tone Segments in the Intonation of Dutch', in T. F. Shannon and J. P. Snapper (eds.), *The Berkeley Conference on Dutch Linguistics 1989* (Lanham MD: University Press of America), 139–55.

—— (1993), 'The Dutch Foot and the Chanted Call', *Journal of Linguistics*, 29: 37–63.

—— (2003), 'Vowel Duration, Syllable Quantity and Stress in Dutch', in K. Hanson and S. Inkelas (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky* (Cambridge, MA: MIT Press, also ROA 381).

——, and BRUCE, G. (1999), 'Word Prosody and Intonation', in H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (Berlin: Mouton de Gruyter), 233–71.

——, and RIETVELD, T. (1992). 'A Target-interpolation Model for the Intonation of Dutch', *ICSLP2*, 1235–8.

——, —— (2000), 'The Behavior of H* and L* under Variations in Pitch Range in Dutch Rising Contours', *Language and Speech*, 43: 183–203.

HALLIDAY, M. A. K. (1970), *Intonation* (London: Oxford University Press).

HART, J. 'T (1998), 'Intonation in Dutch', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems. A Survey of Twenty Languages* (Cambridge: Cambridge University Press), 96–111.

——, COLLIER, R., and COHEN, A. (1990), *A Perceptual Study of Intonation: An Experimental Phonetic Approach* (Cambridge: Cambridge University Press).

HAYES, B. (1995), *Metrical Stress Theory: Principles and Case Studies* (Chicago: University of Chicago Press).

——, and LAHIRI, A. (1991), 'Bengali Intonational Phonology', *Natural Language and Linguistic Theory*, 9: 47–96.

KOHLER, K. J. (1990), 'Macro and Micro Fo in the Synthesis of Intonation', in J. Kingston and M. E. Beckman (eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press), 115–38.

LADD, D. R. (1980), *The Structure of Intonational Meaning: Evidence From English* (Bloomington, IN: Indiana University Press).

—— (1983), 'Phonological Features of Intonational Peaks', *Language*, 59: 721–59.

—— (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

NOLAN, F., and GRABE, E. (1997) 'Can ToBI Transcribe Intonational Variation in the British Isles?', in A. Botinis, G. Kouroupetroglou, and G. Caryannis (eds.), *Intonation: Theory, Models and Application, Proceedings ESCA Workshop*. Athens: ESCA and University of Athens.

O'CONNOR, J. D., and ARNOLD, J. F. (1973), *Intonation of Colloquial English* (London: Longman).

Pierrehumbert, J. (1980), *The Phonetics and Phonology of English Intonation*, Ph.D. dissertation (Massachusetts Institute of Technology), (New York: Garland Press, 1990.)

Post, B. (2000), *Tonal and Phrasal Structures in French Intonation* (The Hague: Thesus, Holland Academic Graphics).

Trim, J. L. M. (1959), 'Major and Minor Tone Groups in English', *Le Maître Phonétique*, 112: 26–9.

Visch, E. (1999), 'Stress Shift', in H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (Berlin: Mouton de Gruyter), 161–232.

Zonneveld, W., Trommelen, M., Jessen, M., Rice, C., Bruce, G., and Árnason, K. (1999), 'Wordstress in West Germanic and North Germanic languages', in H. van der Hulst (ed.), *Word Prosodic Systems in the Languages of Europe* (Berlin: Mouton de Gruyter), 477–603.

6

Transcribing Serbo-Croatian Intonation

Svetlana Godjevac

6.1. INTRODUCTION

The main goal of this chapter is to introduce the intonational system and intonational annotation conventions for Standard Serbo-Croatian.[1] This chapter is organized into two parts. The first part provides a brief introduction to the intonational system of Standard Serbo-Croatian, and the second part is devoted to describing the transcription system worked out within the ToBI family of prosodic transcription systems.

6.2. PROSODIC STRUCTURE OF STANDARD SERBO-CROATIAN

6.2.1. *Descriptive generalizations*

Serbo-Croatian is both a pitch accent language and a stress language. It differs from languages like English, German, Italian, or Dutch, among many others that employ pitch accents for pragmatic highlighting, in that lexical pitch accents are assigned to prominent syllables at the level of the lexicon and thus are used to differentiate word meanings.

Traditional analyses (Lehiste and Ivić 1963, 1986; Brown and McCawley 1965; Nikolić 1970; Gvozdanović 1980; Kostić 1983; Inkelas and Zec 1988;

I would like to thank Mary Beckman, Chris Barker, Allison Blodgett, Mariapaola D'Imperio, Rebecca Herman, Sun-Ah Jun, Julie McGory, Shu-hui Peng, Jennifer Venditti, all the participants of the Summer 1999 class on 'Intonational Analysis and Prosodically Transcribed Corpora' taught by Mary Beckman and Julie McGory, and the participants of the ICPhS 1999 Workshop on 'Intonation: models and ToBI labeling'.

[1] By the term Standard Serbo-Croatian, I here refer to the Eastern Standard Variant.

Stevanović 1989) recognize four different accents: short falling, long falling (marked ["], and [^] respectively), and short rising, and long rising (marked [´], and [`]). According to all of the previous analyses, the pitch accents in Serbo-Croatian are tied to the stressed syllable of a word.[2] The pitch accent is a property of content words. Functional words, such as prepositions, conjunctions, verbal auxiliaries, and pronominal clitics bear no stress or pitch accent.

The distribution of the accents is sensitive to their melody. The descriptive generalization is that the falling accents can occur only on words with the stress on the first syllable, the rising accents can occur on words that have the stress on any syllable but the last. This effectively reduces the distribution of the rising accents to polysyllabic words exclusively. That is, only the falling accents can occur on monosyllables, since the falling accents are not prohibited from occurring on the last syllable. Because the falling accents never occur on any other syllable but the first,[3] the falling/rising opposition is possible only in domains in which the two accents overlap in distribution: polysyllabic words with the stress on the first syllable.

On the phrasal level, previous analyses have recognized three different patterns: (a) a declarative utterance pattern (Lehiste and Ivić 1986; Inkelas and Zec 1988), (b) a prosodic question[4] pattern (Lehiste and Ivić 1986), and (c) the vocative chant (Inkelas and Zec 1988). The declarative pattern is realized as a fall of the pitch at the end of the utterance. The question pattern is realized with a rise of the pitch at the end of the utterance; and the vocative chant is realized as a mid-fall. According to the analyses presented in Lehiste and Ivić (1986) and Inkelas and Zec (1988), all three phrasal patterns affect the lexical pitch accent associated with the word at the end of the utterance. In other words, the phrasal prosodic pattern neutralizes the lexical pitch accent of the word occurring at the end of the phrase. In their autosegmental analysis, Inkelas and Zec provide an account in which the phrasal level tones overwrite the lexical tones (Inkelas and Zec 1988). In the next section I sketch a set of assumptions that guide the prosodic annotation conventions proposed here.

[2] More specifically, these analyses claim that the pitch accent is a property of the mora, which is the tone bearing unit in Serbo-Croatian. However, for our purposes, it will be sufficient to refer only to the stressed syllable as the tone bearing unit.

[3] There are few exceptions involving short-falling accents on a non-initial syllable in compounds, however.

[4] I define a prosodic question as an utterance with the semantic force of a question but a syntax of a declarative statement. That is, a prosodic question is syntactically an indicative statement that gets its interrogative force from the prosodic pattern.

6.2.2. *Intonational phonology of Serbo-Croatian*

In this section I present a novel autosegmental-metrical analysis of lexical and phrasal level tones of Serbo-Croatian. The analysis presented here builds on the previous accounts by adopting the descriptive generalizations about the distribution of lexical pitch accents and the three phrasal patterns mentioned in the previous section. In addition, the analysis is extended to include two more intonation contours: syntactically marked yes-no questions, and the signalling of continuation.

Unlike the other autosegmental analysis of Serbo-Croatian intonation presented in Inkelas and Zec (1988), this analysis assumes sparse rather than full specification of tones at the surface structure. Evidence for sparse specification of tones in Serbo-Croatian is similar to the argument presented in Pierrehumbert and Beckman (1988) for Japanese. It is based on the observation of Fo slopes found in clitic clusters[5] of different lengths. According to the theory of full tonal specification (Inkelas and Zec 1988), clitics are always specified for L tones at the surface. If so, the Fo associated with the clitic cluster is predicted to show an immediate fall after a disyllabic word bearing a rising accent regardless of the number of clitics. Instrumental evidence (Godjevac 2000*a*) shows that clitic clusters of different lengths (one clitic, two clitics, three clitics) have significantly different slopes. The difference in slope is easily explained under the assumption that clitics are not specified for tone at the surface, in which case the slopes are a function of interpolation between two tonal targets that clitics separate.

(i) *Lexical pitch accents*: one of the modifications of the traditional analyses in this proposal involves lexical pitch accents. Even though traditional descriptions of Serbo-Croatian posit four distinct types of pitch accent, they are reducible to two: falling (H*+L) and rising (L*+H) (Godjevac 2000*b*). In this system, the lexical pitch accents are analysed as localized pitch events. They are represented as a bitonal sequence localized around the stressed syllable. That is, the accents are conceived as a bitonal sequence in which only one tone, the starred tone, is anchored to the tone bearing unit, as in English (Pierrehumbert 1980).

The two falling accents realize the falling melody by a sequence of an H tone immediately followed by an L tone, where the H is a starred tone and the L is a trailing tone. This means that the H tone is anchored to the tone bearing unit of the stressed syllable. The two rising accents express a rising

[5] Serbo-Croatian has syntactic clitics (weak pronominals and verbal auxiliaries) which must cluster together in the so-called 'second position' in a clause. The second position is usually defined as 'the position after the first word or after the first syntactic constituent'.

melody realized as a sequence of an L tone immediately followed by an H. The L tone is the starred tone and the H tone is the trailing tone.[6] This analysis of the rising accents differs from the one proposed in Inkelas and Zec (1988) in the tonal characterization of the relevant tone bearing units. Inkelas and Zec analyse the rising accents as a sequence of two H tones, i.e., HH, over the same set of tone bearing units. Since they assume full tonal specification, the 'rise' of the rising accents is represented at the surface level. At the surface, the rest of the tone bearing units are specified for an L tone (by the default L tone insertion) and hence the rising melody is captured.

Figures 6.1 and 6.2 show a schematic pitch track of the two falling accents in sentence medial position. The difference between the long and the short accent is manifested in the realization of the fall: the fall in the long accent occurs within the second half of the stressed syllable—the second mora of the first syllable; in the short falling, on the other hand, the realization of the fall varies with the length of the word and the position of the word within a phrase: in utterance non-final positions, the fall in polysyllabic words is realized within the second syllable, whereas in utterance final positions, and in monosyllabic words, the fall is realized at the end of the stressed syllable, or in some cases it may be even be truncated.

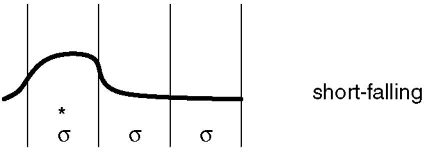


FIGURE 6.1 A schematic contour of the short falling pitch accent in sentence medial position.

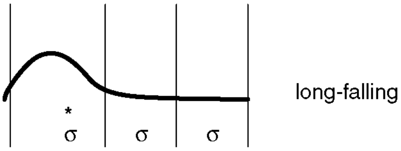


FIGURE 6.2 A schematic contour of the long falling pitch accent in sentence medial position.

Figures 6.3 and 6.4 illustrate the two rising accents. The main difference in the pitch contour between the falling and the rising accents is that in the rising accents the pitch of the post-stressed syllable is higher than that of the stressed syllable, unlike in words under the falling accents. The difference between the

[6] This analysis is also supported by the quantitative study of Smiljanić and Hualde (2000).

long and the short version of the rising accent is only manifested in the length of the stressed syllable. At the level of the syllable, the long vowel is represented as two morae, and the short as one mora (see Zec 1994). The combination of the accent specification with the mora differentiation of syllable length is sufficient to guarantee the four accent types of the traditional analyses.

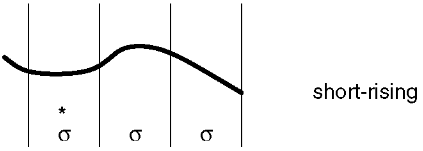


FIGURE 6.3 A schematic contour of the short rising pitch accent in sentence medial position.

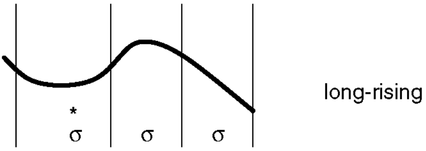


FIGURE 6.4 A schematic contour of the long rising pitch accent in sentence medial position.

(ii) *Phrase accents and boundary tones*: Serbo-Croatian has at least five distinct shapes that occur at ends of phrases. They are interpreted either as: (a) a declarative, (b) a prosodic question, (c) signalling of continuation, (d) a yes-no question, or (e) a vocative chant (calling contour). Three of these contours (a, b, and e) have already been acknowledged by the previous analyses. The remaining two are introduced here.

Of the three shapes recognized by the traditional analyses, only the vocative chant is here given a comparable analysis to the one originally proposed in Inkelas and Zec (1988), i.e., it is treated as a bitonal boundary tone. The other two shapes, the declarative fall, and the prosodic question, have been re-analysed as complex shapes consisting of a phrase accent and a boundary tone. The shape that we find in yes-no questions is also analysed as a complex shape composed of a phrase accent and a boundary tone, whereas the continuation rise is analysed as an H boundary tone.

The assumption in the analysis of the phrasal Fo shapes adopted in this chapter is that phrase accents are tonal targets that are linked to metrically strong positions. Boundary tones, on the other hand, are tonal targets occurring at the phrase's edge (Pierrehumbert 1980; Grice *et al.* 2000). This means that phrase accents align with respect to stressed syllables, and boundary tones are realized over syllables occurring at the edge of a phrase, i.e., either around the first or around the last syllable. The interaction of lexical level tones and phrasal tones is one of the major topics in the study of Serbo-Croatian prosody. Given that phrase accents target metrically strong positions just as lexical pitch accents do, phrase accents often neutralize the lexical level tones. Boundary tones, on the other hand, are not in competition for the same targets as lexical level tones, although in some cases (monosyllabic and disyllabic words for example) they do coincide. Consequently, boundary tones do not overwrite lexical tones, as phrase accents do, although they may override them in certain contexts.

The proposed inventory of phrase-level tones in Serbo-Croatian consists of two phrase accents, LH-, and Ø-, and five boundary tones: final boundary tones: L%, H%, word initial boundary tones: %L, %H, and a bitonal boundary tone: HL%. In what follows, I discuss the properties of phrasal tones in Serbo-Croatian, starting with intonational phrase boundary tones.

(a) *Intonational phrase boundary tones*: there are two simple boundary tones: H% and L%, and one complex boundary tone, HL%. The two simple boundary tones also combine with phrase accents.

The L% tone signals finality, whereas the H% tone signals the absence of finality (continuation or questioning). Since these tones occur on the last syllable of the last word within the intonational phrase, they do not neutralize the lexical pitch accents, although they modify their realizations. In the previous section we have seen a schematic representation of the lexical pitch accents in phrase medial position. When delimited by a L% boundary tone, the Fo of words under the falling accents keeps falling, turning into a creaky voice, whereas the Fo of words under the rising accents stays level, unlike in sentence medial positions where the post-stressed syllable shows a clear rise. The effect and the realization of the L% boundary tone in these two environments is schematically represented in Figures 6.5 and 6.6. Figures 6.7 and 6.8 show schematic representation of the H% boundary tone under the different lexical accents.

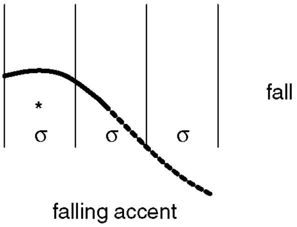


FIGURE 6.5 A schematic representation of the L% tone realized by a word with a lexically specified falling accent.

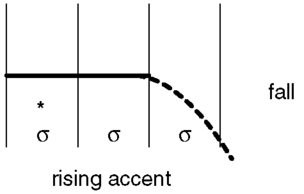


Figure 6.6 A schematic representation of the L% tone realized by a word with a lexically specified rising accent.

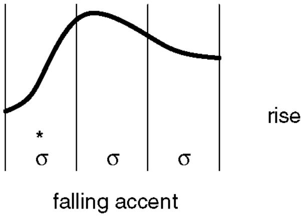


Figure 6.7 A schematic representation of the H% tone realized by a word with a lexically specified falling accent.

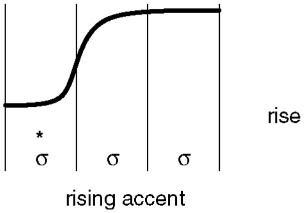


Figure 6.8 A schematic representation of the H% tone realized by a word with a lexically specified rising accent.

The third type of boundary tone is a bitonal sequence, HL, that creates a mid fall. The HL% boundary tone is a calling contour and it spans the last two syllables of the word. A schematic representation of the Fo shape of this boundary tone is shown in Figure 6.9.

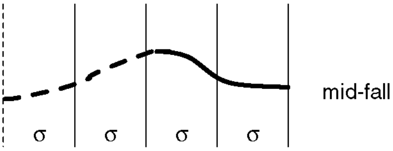


Figure 6.9 A schematic contour of the HL% boundary tone.

(b) *Phrase accents*: the Ø- phrase accent is a property of the nuclear position within an intonational phrase in declaratives, *wh*-questions, and imperatives. All of these types of utterances are delimited by a L% boundary tone. This phrase accent is associated with the focused constituent in the case of narrow focus and the rightmost constituent in the intonational phrase in the case of broad focus. Its distinguishing property is not a tonal target, but a compression of the pitch range for the material following it. This is shown in Figure 6.10.

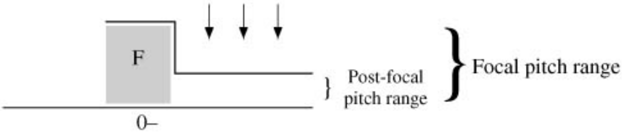


Figure 6.10 A schematic representation of the effect of a Ø- phrase accent.

The LH- phrase accent occurs on the focused constituent in prosodic questions and syntactically marked yes-no questions. The two types of questions differ in the type of boundary tone that follows. Prosodic questions are delimited by an H% boundary tone, whereas syntactically marked yes-no questions are delimited by an L% boundary tone. The pragmatic force of the phrase accent is to convey a sense of surprise and questioning for the focused word.

In syntactically marked yes-no questions, this phrase accent aligns to the stressed syllable of the focused word (usually the verb[7]). A schematic illustration of this accent in yes-no questions is given in Figures 6.11 and 6.12. The shape of this accent is identical for both the falling and the rising accents, however, the alignment properties differ: the falling accents realize the peak earlier than the rising accents. The falling accents realize the peak at the end of the stressed syllable, whereas the rising accents realize the peak at the end of the post-stressed syllable. This means that the distinction between the lexical pitch accents under this phrase accent is maintained.

[7] In yes-no questions this accent is realized on the verb by default, since the verb is associated with the meaning of polarity. However, the accent may occur on any other constituent as well, if the focus of the question is something other than the polarity itself.

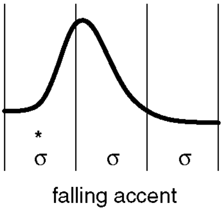


FIGURE 6.11 A schematic contour of the LH- phrase accent in yes-no questions for a trisyllabic word under the falling accent.

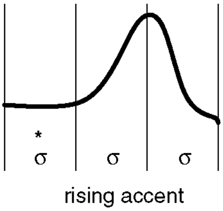


FIGURE 6.12. A schematic contour of the LH- phrase accent in yes-no questions for a trisyllabic word under the rising accent.

In prosodic questions, the LH- phrase accent must occur on a focused word. In this case there is no default placement. Because this type of question is not marked syntactically, the phrase accent must be followed by a H% boundary tone in order to signal the interrogative force of the utterance.[8] The pragmatic function of this phrase accent in combination with the H% boundary tone is in many respects similar to the incredulity question contour in English, i.e., L*+H L- H%, discussed by Ward and Hirschberg (1988). The similarity lies in the interpretation of the prosodic pattern, i.e., both contours convey incredulity.

When the LH- phrase accent is followed by a H% boundary tone, the alignment properties of these two tones are different from their counterparts in the context of the L% boundary tone. Here, the H tone targets the final syllable, rather than the post-stressed syllable of the focused word. The alignment of the L tone is the same, the L tone aligns with the stressed syllable. This alignment is identical for both pitch accent types and hence the lexical pitch accent distinctions are lost in this environment. A schematic

[8] Lehiste and Ivić call this melody a 'reverse pattern'. They call it a 'reverse pattern' because in this contour the stressed syllable is associated with the Fo minima and the post-stressed syllable(s) with the Fo maxima. They view this as a reversal of the usual pattern for signalling prominence where the stressed syllable is the one associated with the peak. Consequently, they term the Fo minima in this prosodic pattern a 'negative peak'.

illustration of the combination of a LH- phrase accent and a H% boundary tone for the two types of lexical pitch accents is shown in Figures 6.13 and 6.14.

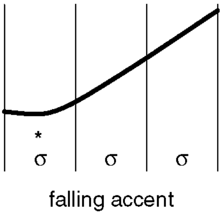


Figure 6.13 A schematic contour of the LH- phrase accent in prosodic questions (i.e., preceding a H% tone) for a trisyllabic word under the falling accent.

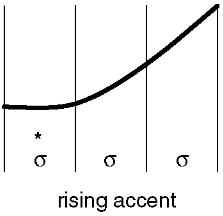


Figure 6.14 A schematic contour of the LH- phrase accent in prosodic questions (i.e., preceding a H% tone) for a trisyllabic word under the rising accent.

(c) *Word boundary tone*: there are two types of word boundary tones: a %L and a %H. In broad focus utterances, all accented words and some non-accented words (prepositions for example) are delineated by an initial %L word boundary tone. The %L word boundary tone is illustrated in the next section where prosodic constituents are discussed. The %H boundary tone marks a word as a contrastive or corrective narrow focus and it raises the pitch range for the focused word.

The two word boundary tones occur word initially. The evidence for the left edge attachment of a word boundary tone as opposed to the right is twofold. The first piece of evidence comes from utterance initial positions which show the presence of some Fo minima. These can be interpreted as the L tone target word initially (see Figure 6.15). The second piece of evidence comes from the Fo contours found on proclitics. When a proclitic is attached to a word under the falling accent, which means that the first tone target is an H, we never see a gradual slope towards the accent H starting from the proclitic, but rather we find a steady flat Fo up to the beginning of the word and then a steep rise towards the lexical H. This shows that there is a clearly marked point at which the rise starts. Since this point always occurs at the beginning of the word we can explain it by positing a %L word boundary tone at the left edge of the word.

(iii) *Prosodic constituents*: Serbo-Croatian has two prosodic constituents that are defined intonationally: the phonological word, and the intonational phrase. The phonological word is defined by two tonal events: the pitch accent, and the initial %L (or %H when in focus) word boundary tone.

The intonational phrase is defined by the right edge boundary tone, a phrase accent, phrase-final lengthening, and a pause. The intonational phrase also defines a domain for the local pitch range computation, i.e., downstep. The best introduction to the two types of prosodic groupings is by example. Figure 6.15 shows a pitch track of a broad focus declarative utterance which consists of a single intonational phrase that contains five phonological words. The phonological words are delineated by a %L word boundary tone, indicated by an arrow in the figure. The prosodic constituency of this utterance is shown in (1).

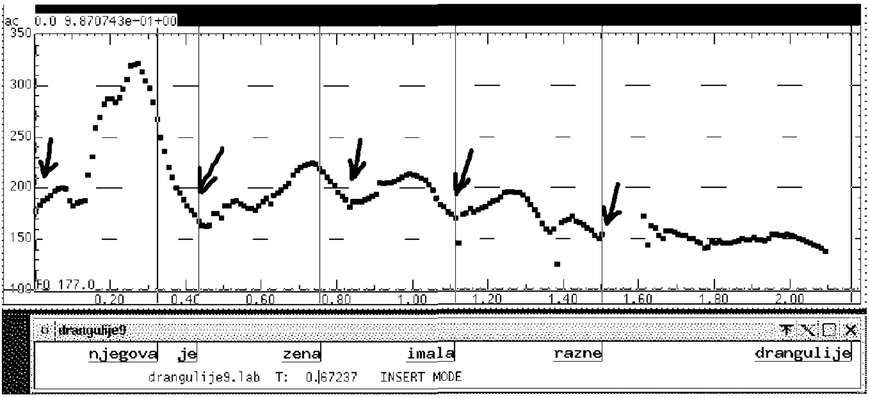


FIGURE 6.15 An Fo track of the sentence *Njegova je žena imala razne drangulije*, 'His wife had all sorts of junk,' the arrows indicate the point in the Fo where a L word boundary tone is posited.

(1)	Njegova	je	žena	imala	razne	drangulije.	
	His	aux	wife	had	all sorts	junk	
	(	)	( )	( )	( )	(	) Phonological words
	[						] Intonational phrase

The pitch range of each phonological word within the intonational phrase is smaller than that of the preceding one. In other words, the pitch range is downstepped for each subsequent phonological word. This is evident in Figure 6.15. However, when an intonational phrase contains more than five phonological words, there is usually an adjustment of the downstepping of the pitch range. This adjustment consists of breaking the downtrend and locally raising the pitch range to the same level of or slightly higher than the preceding word in order for the downtrend to be able to sustain the remaining text within the intonational phrase.[9] This pitch range adjustment is a function of the length of the utterance. The choice for the placement of this adjustment seems to be free within the intonational phrase.[10] The pitch range adjustment is illustrated in Figure 6.16. The arrows indicate the point at which the readjustment occurred: the peak at the point of the pitch range readjustment is higher than or equal to the preceding one.

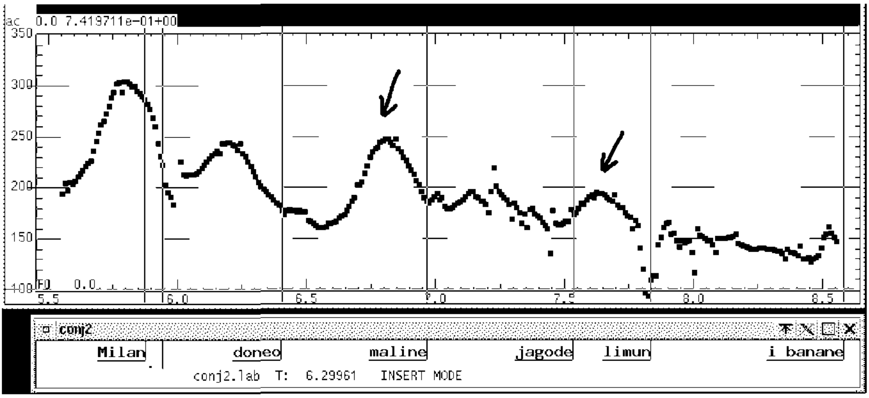


Figure 6.16 An Fo track of the sentence *Milan je doneo maline, jagode, limun i banane* 'Milan brought raspberries, strawberries, lemon, and bananas.'

In this section we have seen some major characteristics of the Serbo-Croatian prosodic system. In order to build a more comprehensive understanding of Serbo-Croatian intonation we need a larger database of spoken utterances of all types and registers. Having a system for annotating spoken utterances for their prosodic properties is of considerable importance in

[9] Failure to adjust the pitch range in longer utterances signals a very monotone, bored speech.

[10] This phenomenon needs to be carefully annotated in order to study it further. One possible interpretation of the pitch range adjustment is another level of grouping within the intonational phrase. Although I did not have enough evidence to support this proposal, this hypothesis should be carefully reconsidered in a larger data set.

creating useful databases that can serve as a tool for testing and refining hypotheses about the intonational system. The next section is devoted to an annotation system designed for Serbo-Croatian based on the intonational system presented in this section. This system is far from complete, however, I hope that even as such it can provide us with useful annotation that would allow for the testing of hypotheses that are not necessarily compatible with this system.

6.3. ToBI SYSTEM FOR STANDARD SERBO-CROATIAN

The annotation system introduced here is modelled after other more established ToBI framework systems, such as English (see Beckman and Hirschberg 1993), and Japanese ToBI (see Venditti 1995). The aim of this system is an accurate symbolic description of Fo contours of diverse speech types, with utterance styles ranging from read to spontaneous. The ultimate goal, however, is to use this system to transcribe other dialects of Serbo-Croatian not covered in the phonological analysis and thus facilitate interdialectal comparisons. I believe that this can easily be accomplished by slight modifications and/or amendments to the present system.

The SC-ToBI system currently includes a wave form, an Fo curve, and five separate label tiers. At this stage a spectrogram is not included, but it can easily be accommodated. The labelling tiers are devoted to descriptions of (1) tones, (2) words, (3) break indices, (4) glosses, and (5) miscellaneous non-linguistic information. In the following sections each tier is described and discussed separately with illustrations from the current database. For the moment, this database consists of read speech style only.

6.3.1. *Word tier*

The word tier contains orthographic marking of Serbo-Croatian words. Words are taken to be morpho-syntactic units as defined by the conventions of the writing system of Serbo-Croatian. They are marked at their right edge according to their waveform (and/or spectrogram).

Serbo-Croatian orthography is very close to phonemic and few changes need to be made. We use standard romanization with a few adaptations required to make the labelling conform to ASCII characters. Table 6.1 includes the relevant orthographic changes.

Table 6.1 Adaptation of Serbo-Croatian alphabet to ASCII format

č = c^	ć = c'	š = s^	ž = z^

Even though vowel length is contrastive, it is not marked in the orthography. However, since this distinction is crucial for differentiating short and long accents, vowel length is marked in the word tier. That is, all long syllables, stressed or not, are marked as 'V:'. Thus, the conventional spelling of a word under a long accent, such as *rad* 'work', in this system will be spelled as *ra:d.* Since the tone tier only marks accents with respect to falling/rising opposition, the word tier remedies this deficiency.

6.3.2. *Tone tier*

(i) *Lexical pitch accents*: the falling accents are to be marked at their peak, if possible, otherwise at the middle of the stressed syllable. The rising accents are to be marked at the lowest Fo value in the stressed syllable if possible, otherwise in the middle of the stressed syllable. Reasons for not being able to mark the peaks or valleys of these two accent types include pitch range compression, common to post-focal stretches due to the Ø- phrase accent. We also find delayed peaks, because of the word boundary tone. The delayed peaks are to be marked '>'. The next two figures illustrate markings of the lexical pitch accents.

Figure 6.17 shows pitch tracks of two phrases consisting of words with rising accents only. The two sentences provide examples of both types of rising accents (short and long). The most representative Fo shape of the two types of rising accents is in utterance medial position, the word *línija* 'line' (long-rising) in the left side of the panel, and the word *màlina* 'raspberry' (short-rising) in the right side of the panel. The tone tier does not provide different labels for the short/long distinction of the rising accents. However, as Figure 6.17 shows, they are kept apart in the word tier by differentiating long vowels from short ones using the standard IPA symbol for vowel length. Figure 6.18 gives Fo contours of two utterances which contain examples of the two falling accents in sentence medial and final position.

Marking lexical accents is fairly straightforward in broad focus utterances such as those found in Figures 6.17 and 6.18, because the shape of the Fo contour provides an easy visual differentiation between the two types of

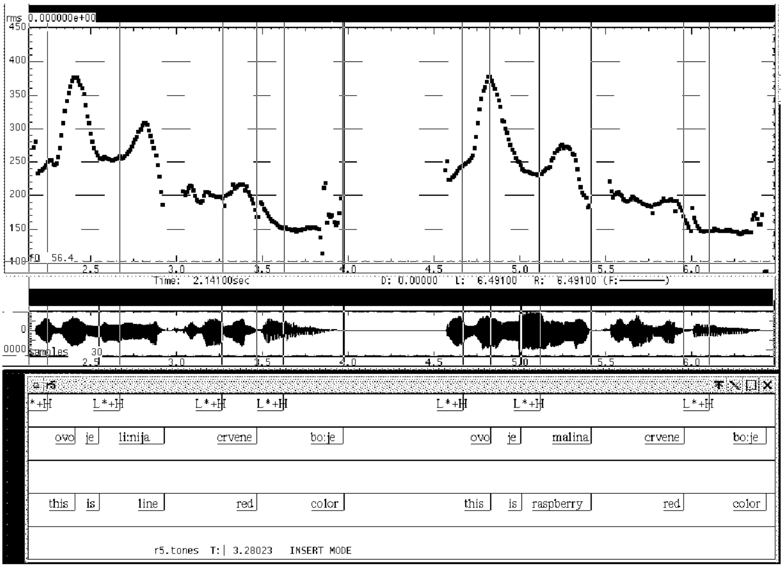


Figure 6.17 An example of an utterance with all words under the rising accents. The best illustration of the two types of rising accents is provided by the words in utterance medial positions: *línija* 'line' for the first utterance and *màlina* 'raspberry' for the second. Gloss: *Ovo je línija crvene boje. Ovo je màlina crvene boje.* 'This is a red-coloured line. This is a red-coloured raspberry.'

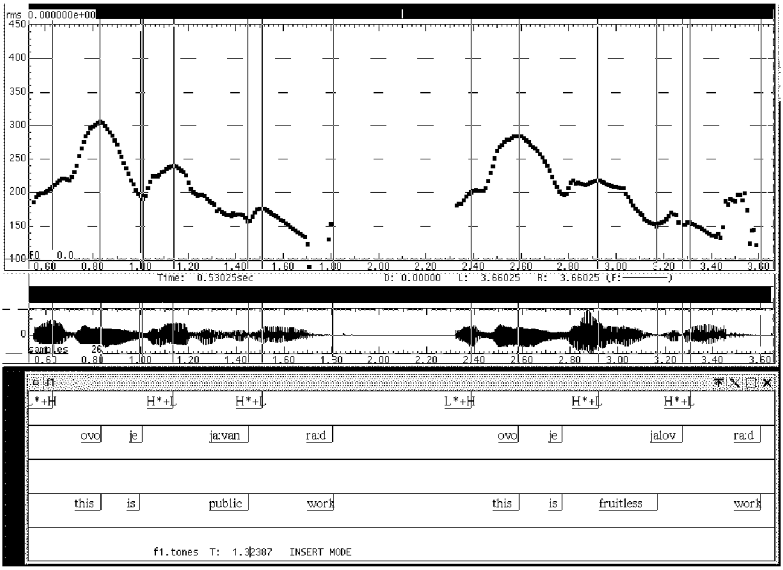
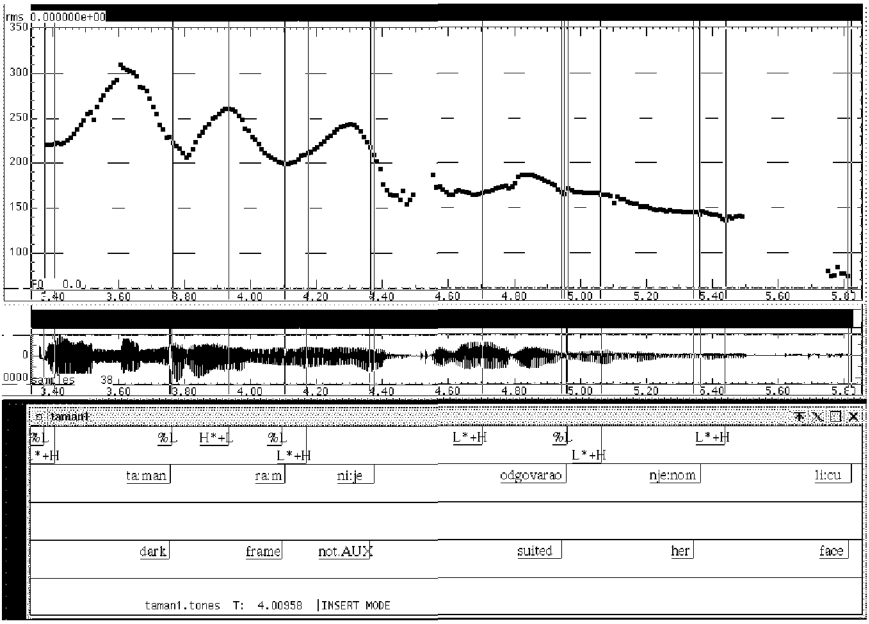


Figure 6.18 An example of an utterance with falling accents in utterance medial positions. Gloss: *Ovo je jâvan rad. Ovo je jȁlov rad.* 'This is public work. This is fruitless work.'

accents. However, higher level markings associated with narrow focus often obscure the lexical tones. The label '*?' is reserved for those cases where it is not possible to tell if a pitch accent is present, and the label 'X*?' is to be used when the nature of the pitch accent is at issue.

(ii) *Word boundary tone*: as was mentioned earlier, there are two types of boundary tones that occur at the left edge of a phonological word: a %L and a %H word boundary tone. The %L tone occurs in broad focus utterances and the %H boundary tone is used to signal narrow prosodic focus of a contrastive or corrective focus. These tones are marked at the left edge of the phonological word to which they belong. Figure 6.19 provides a number of illustrations of the %L boundary tone, whereas Figure 6.20 provides an example of the %H boundary tone on the focused word. Word boundary tones are marked at the beginning of the word.



```
taman1.tones  T:  4.00958  |INSERT MODE
```

FIGURE 6.19 An illustration of the %L word boundary tone. Gloss: *Taman ram nije odgovarao njenom licu.* 'A dark frame didn't suit her face.'

(iii) *Intonational phrase boundary tones*: there are three types of boundary tones that can delimit an intonational phrase in Serbo-Croatian: L%, H%, and HL%. The L% occurs with phrases that are syntactically marked for their semantic force (declaratives, interrogatives, imperatives) and it aligns with the last syllable in the intonational phrase. The H% delimits and marks an

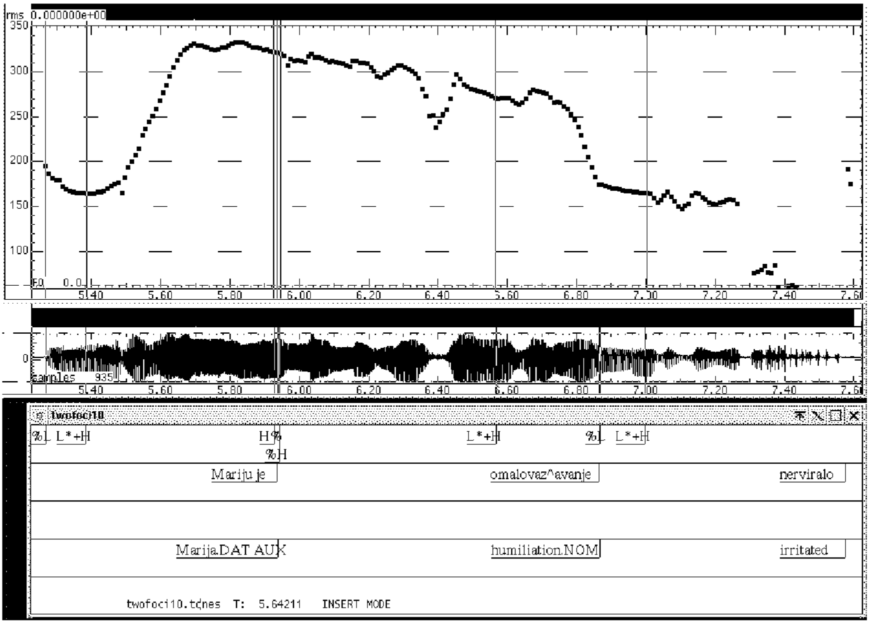


FIGURE 6.20 An illustration of the %H word boundary tone. Gloss: *MARIJU je OMALOVAŽAVANJE* nerviralo. 'MARY was irritated by the HUMILIATION.'

intonational phrase as a question or a phrase that signals the expectation of a continuation. This boundary tone occurs with phrases which are syntactically unmarked as questions or with phrases that are sentence fragments, and aligns with the last syllable within the phrase. The HL% boundary is the vocative chant.

All boundary tones are marked at the right edge of the intonational phrase. The L% boundary tone is present at the end of utterances in all previous figures. This boundary tone in association with the Ø phrase accent is typical with declaratives, *wh*-questions, syntactically marked yes-no questions, and imperatives.

A H% boundary tone occurs in utterances uttered as multiple phrases and its function is to signal continuation. Figure 6.21 illustrates an H% boundary tone after the first two words in a sentence. This utterance was uttered by a Serbo-Croatian language teacher in the US, who is used to uttering Serbo-Croatian sentences very slowly and carefully which leads into separation of words into intonational phrases. The use of the H% boundary tone is necessary to signal the absence of completion of the sentence. This speaking style is not common in colloquial speech, but the example illustrates very clearly the continuation rise contour. The right edge of the word *Jèlena* 'Jelena' but also the word *daje* 'gives' carry a H% boundary tone. This utterance thus consists of three intonational phrases. The calling contour, or the HL% boundary tone is illustrated in Figure 6.22. The boundary tone is marked at the right edge of the phrase.

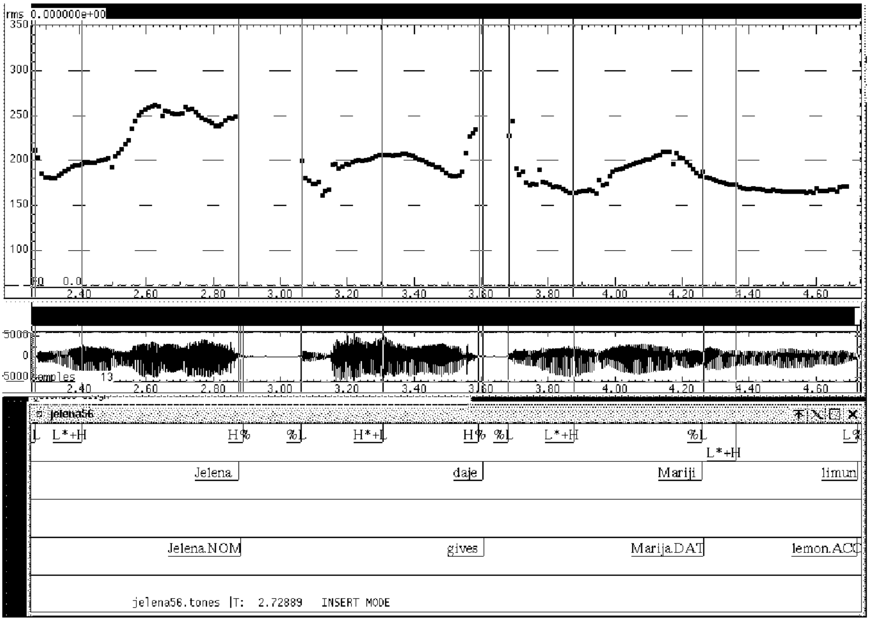


FIGURE 6.21 An illustration of the H% boundary tone. The H% boundary tone in this utterance occurs at the left edge of the word *Jèlena* 'Jelena'. Gloss: *Jelena daje Mariji limun.* 'Jelena is giving Mary a/the lemon.'

(iv) *Phrase accents*: as we have seen earlier, there are two phrase accents, Ø- and LH-. In this section we will see their Fo realizations and the proposed conventions for labelling them.

The Ø- co-occurs with the L% boundary tone in statements, *wh*-questions, and imperatives. It is realized on the word bearing the sentence stress. It does not have a tone target, rather its function is to compress the pitch range for the material following the focus. The effect of this phrase accent is hard to see in broad focus utterances, however, it is strikingly obvious in narrow focus utterances with an early focus. An example of this latter types is shown in Figure 6.23, where there is an early focus and the pitch range compression is quite obvious.

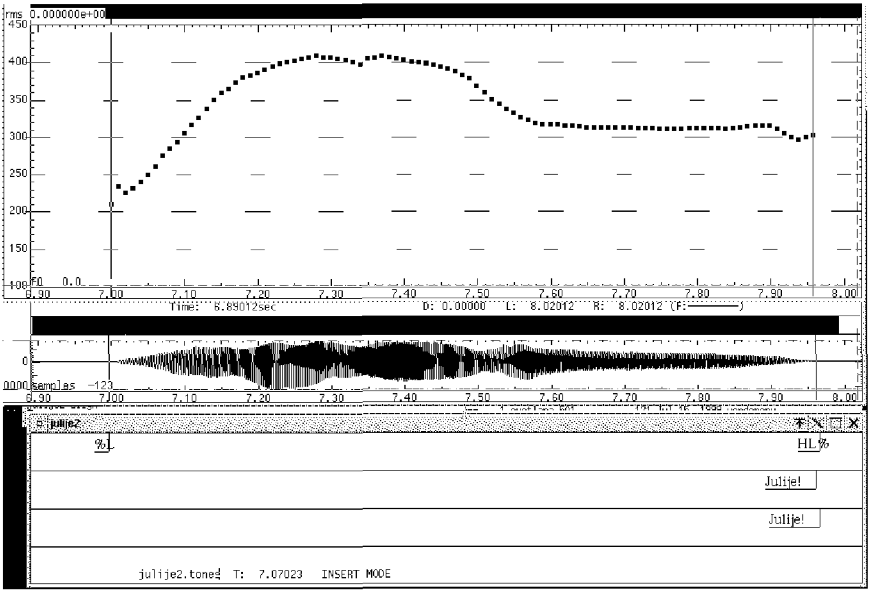


FIGURE 6.22 An illustration of the HL% intonational phrase boundary tone used in vocative chants. Gloss: *Jûlije!* 'Julije!'

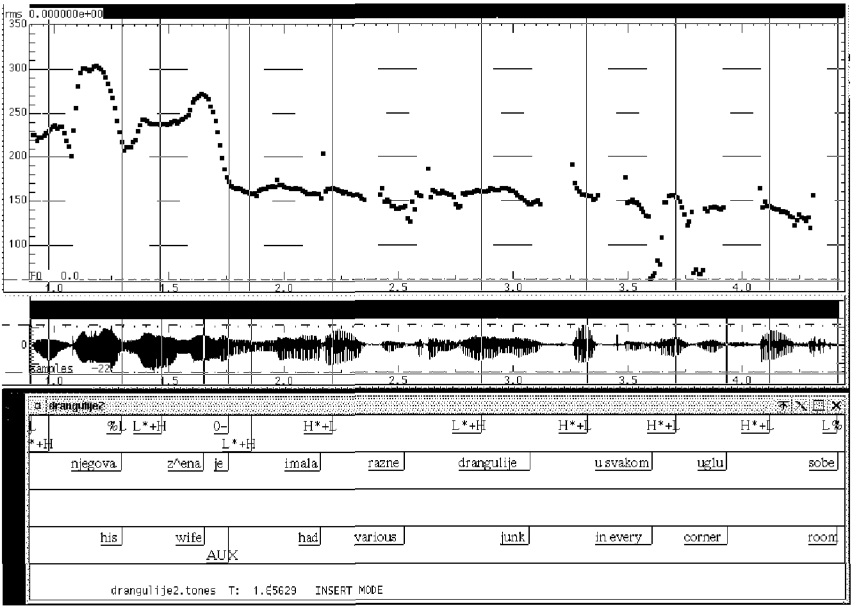


FIGURE 6.23 An illustration of a phrase non-final focus. The focused phrase is *žena* 'wife'. The utterance is: *Njegova žena je imala razne drangulije u svakom uglu sobe.* 'His wife had all sorts of junk in every corner of the room.'

The LH- phrase accent is a property of the focused word in yes-no questions. As we have discussed earlier, the L tone is always anchored to the stressed syllable of the focused word and the H tone is anchored to the last syllable of the same word. In prosodic questions, this phrase accent occurs as close as possible to the right edge of the intonational phrase, possibly because it is tied to the H% boundary tone which raises and compresses the pitch range. The LH- phrase accent is marked at the middle of the stressed syllable of the focused word, which is the anchor for the L tone. The prosodic question contour, i.e., LH-H%, is very common with single word utterances, but it occurs with longer phrases as well. Figure 6.24 illustrates the LH- phrase accent in combination with the H% boundary tone on the word *Marija*, which is consequently the focus of the prosodic question.

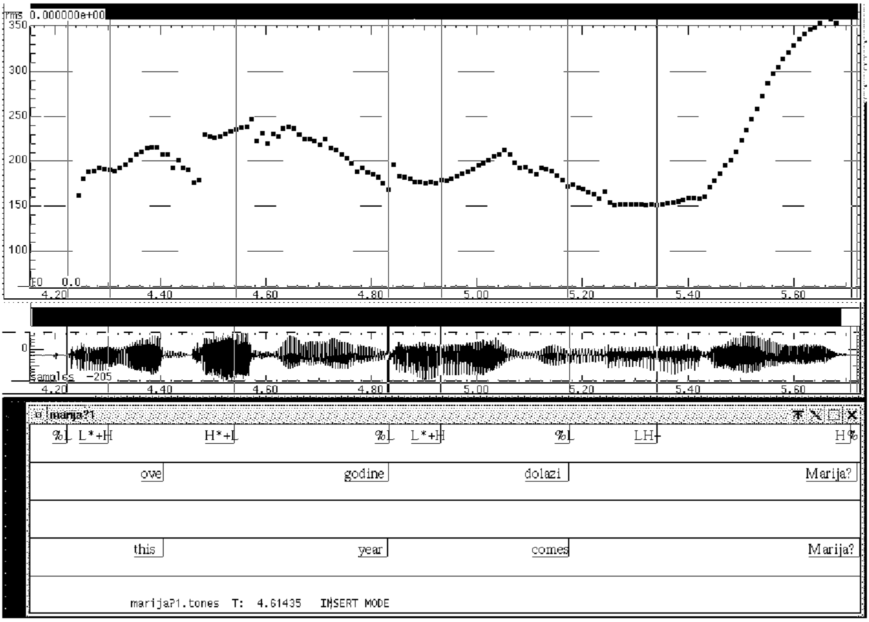


FIGURE 6.24 An illustration of the LH- phrase accent in the utterance: *Ove godine dolazi MARIJA?* (gloss: this year comes Marija) 'MARIJA is coming this year?' The word *Marija* bears the phrase accent and is consequently the focus of the sentence.

Focusing a penultimate word in a sentence in a prosodic question shows the distinction between the LH- phrase accent and the H% boundary tone. This is illustrated in Figure 6.25. In this figure, LH- phrase accent occurs on *ove* 'this', the penultimate phonological word in the sentence, and the rest of the phrase, the word *gòdine* 'year' is in the raised and compressed pitch range.

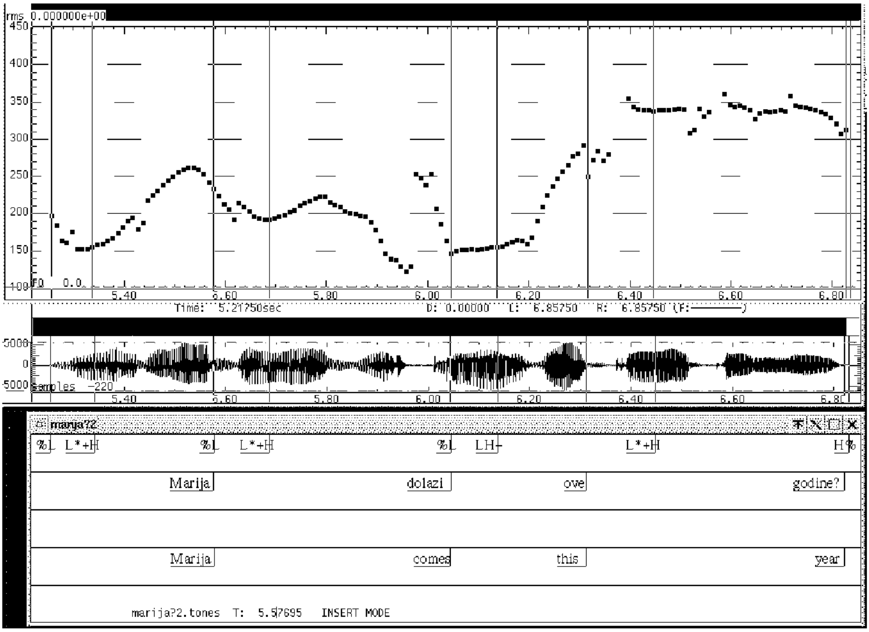


FIGURE 6.25 An illustration of the LH- phrase accent on the penultimate word *ove* 'this' in the utterance *Marija dolazi OVE godine? 'Marija is coming THIS year?'*

In syntactically marked yes-no questions, the LH- phrase accent usually occurs on the verb,[11] and is followed by a L% boundary tone. However, the accent can occur on any of the constituents within the sentence. Figure 6.26 provides an illustration of this accent when occurring on a non-verbal constituent.

(v) *Additional labels*: there are a few additional labels reserved for the tone tier. Delayed Fo is marked '>'. This label is used to mark delayed peaks that we find on the falling accents and delayed lows that are often found with the word boundary tone. The labels 'X%?' and '%X?' are to be used for a boundary tone the labeller is uncertain of. The label '%?' is to be used when the labeller is not quite certain of the presence of a boundary tone. As I mentioned earlier, we also want to mark the pitch range adjustments found in longer utterances. This will be useful in testing the hypothesis about additional prosodic grouping in Serbo-Croatian. The label '#' is used to mark raising of the pitch range. The use of this label is illustrated in Figure 6.27.

[11] The verb is the default placement of the accent because the verb carries the polarity meaning. However, placing prominence on another constituent is also an option.

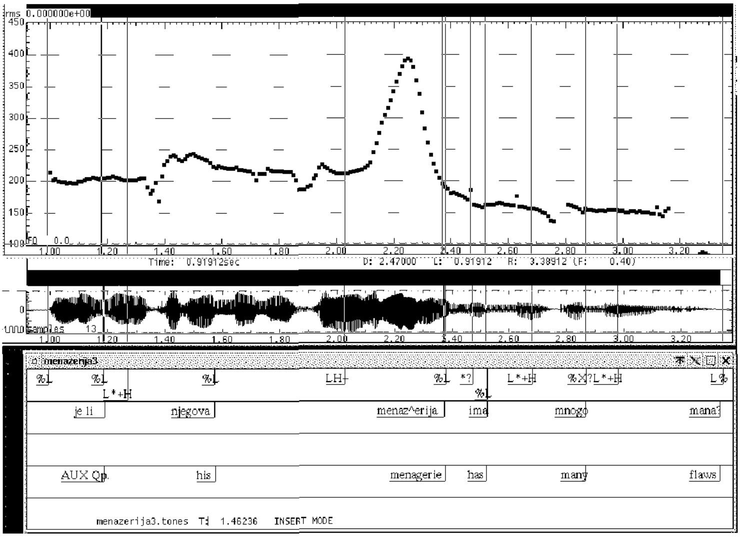


FIGURE 6.26 An illustration of the LH- phrase accent often found on focused constituents in morphologically marked yes-no questions. This pitch track represents the following utterance: *Je li njegova MENAŽERIJA ima mnogo mana?* 'Is it his MENAGERIE that has many flaws?'

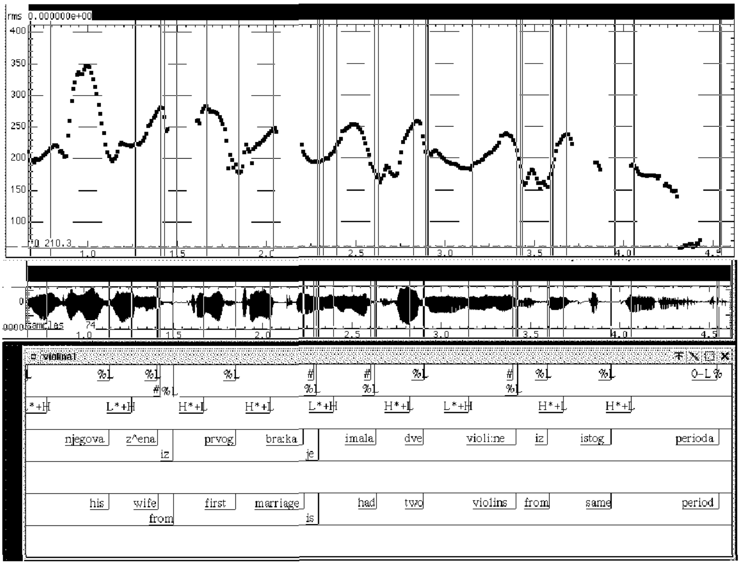


FIGURE 6.27 For ease of readability this panel contains two tone tiers: one for lexical tones (i.e., the pitch accents) and the other for the phrasal tones. The utterance is: *Njegova žena iz prvog braka je imala dve violine iz istog perioda.* 'His first wife had two violins from the same period.'

6.3.3. *Break index tier*

The break index tier is an indicator of the strength of prosodic relations between words. Three levels are proposed: 0, 1, and 2. The break index ‘0’ is an index mark of the strongest bond, as typically found among clitics and their hosts. This prosodic domain is also the one where we find segmental sandhi phenomena, such as regressive devoicing, consonant deletion, etc. The index marker ‘1’, is used for the relation between phonological words. The break index ‘2’ is used for the juncture between two intonational phrases separated by an intonational phrase boundary tone. No segmental sandhi are expected across this juncture.

Typically, break indices match tonal markings of prosodic structure. The break index ‘0’ entails no tonal marking between words. The break index ‘1’ entails a %L word boundary tone, found between words. The break index ‘2’ entails an intonational phrase boundary tone. However, the relation between break indices and tonal markings is not reciprocal. The presence of a particular tonal marking does not necessarily entail a particular break index. For example, we usually find a %L word boundary tone between a proclitic and its host although their bond is very strong and so we mark it as the weakest disjuncture, i.e., ‘0’.

It is also not uncommon to find mismatches between the degree of disjuncture and the tonally defined prosodic domain. A diacritic ‘m’ is used for the presence of a mismatch. In Figure 6.28, there is a mismatch after the first word in the utterance *ono* ‘that’, which is labelled ‘1m’. This label says that the intuitive sense of the boundary found is a ‘1’, however, the tonal structure does not support this boundary since the word boundary tone is not present. Label ‘X’ is used for uncertainty of a break index. This provision is made for those cases that are quite difficult and consequently require more study.

6.3.4. *Gloss tier*

The gloss tier provides English glosses for Serbo-Croatian words. This tier is optional and its use should be at the discretion of the labeller and the purpose of a database. Since most of the research is being conducted in an international community, it seems desirable to facilitate exchanges among scholars by providing easy access to translation.

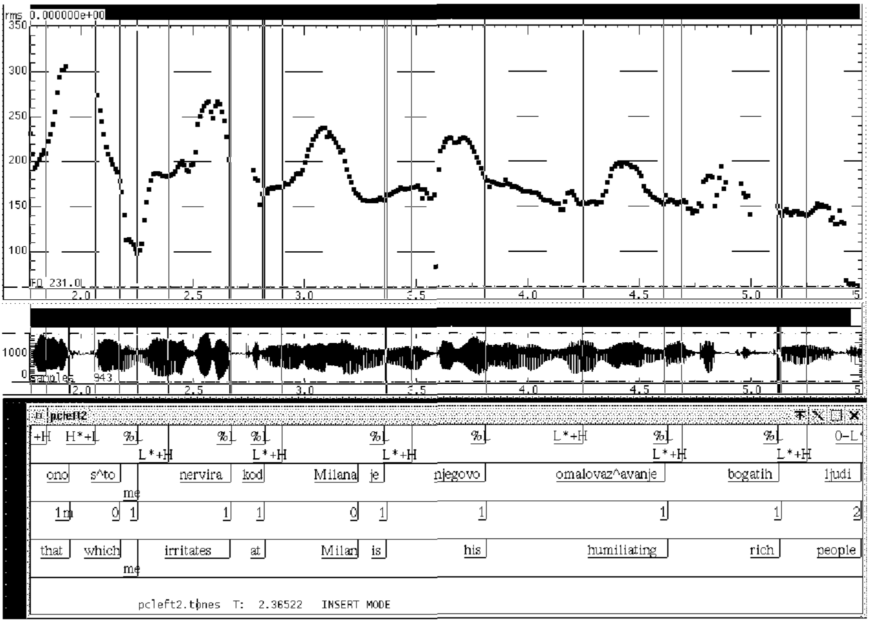

rms 0.000000e+00
350
300
250
200
150
100
F0 231.0L
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5
1000
0
samples 943
pcleft2
+H H*+L %L L*+H %L %L L*+H %L L*+H %L L*+H %L L*+H %L L*+H 0-L
ono s^to me nervira kod Milana je njegovo omalovaz^avanje bogatih ljudi
1m 0 1 1 1 0 1 1 1 1 2
that which me irritates at Milan is his humiliating rich people
pcleft2.tones T: 2.36522 INSERT MODE

FIGURE 6.28 An illustration of the break indices and a mismatch between a break index and a tonal marking between the two first words. The utterance is: *Ono što me nervira kod Milana je njegovo omalovažavanje bogatih ljudi.* 'What irritates me about Milan is his humiliation of rich people.'

6.3.5. *Miscellaneous tier*

The miscellaneous tier is reserved for non-phonological phenomena found in spontaneous speech, such as repairs, disfluencies, coughing, etc.

6.4. SUMMARY

In this chapter we have seen some basic intonational patterns of Standard Serbo-Croatian. Intonational structure of Serbo-Croatian is composed of lexical pitch accents and higher level prosodic markers: phrase accents and boundary tones. Based on this prosodic analysis, a model of prosodic transcription within the ToBI family of annotation has been introduced. A summary of the proposed annotations can be found in Table 6.2.

TABLE 6.2 Summary of SC_ToBI labels

H*+L	*Falling Accents*: marked on the stressed syllable of words with lexical falling accents.
L*+H	*Rising Accents*: marked on the stressed syllable of words with lexical falling accents.
LH-	*Phrase accent*: marked on the stressed syllable of the focused word at the phrase edge signalling surprised questions.
Ø-	*Zero phrase accent*: marked on the last syllable of the word at the phrase edge.
%L	*Low word boundary tone*: marked at the left edge of the phrase.
%H	*High word boundary tone*: marked at the left edge of the focused phonological word.
L%	*Low intonational phrase boundary tone*: marked at the right edge of an intonational phrase.
H%	*High intonational phrase boundary tone*: marked at the right edge of the intonational phrase signalling continuations or questions.
HL%	*Mid-fall boundary tone*: marked at the right edge of the phrase, signalling vocative chants.
0	*Break index: strongest cohesion*: indicates lack of word boundary tone. Typical in host-enclitic junctures.
1	*Break index: strong cohesion*: indicates presence of low word boundary tone. Typical between words and in proclitic-host junctures.
2	*Break index: weak cohesion*: indicates presence of intonational phrase juncture.
*?	*Accent uncertainty*: this mark is reserved for word accents the labeller is unsure of.
>	*Delay*: indicates delayed peak or a delayed L word boundary tone.
%X?	*Word boundary uncertainty*: this mark is reserved for word accents the labeller is unsure of.
X%?	*Intonational phrase boundary uncertainty*: this mark is reserved for intonational phrase boundary the labeller is unsure of.
M	*Mismatch diacritic*: indicates a mismatch between a break index and tonal evidence for it.
#	*Pitch range adjustment*: this mark is used to tag the local pitch range readjustment. It is placed at the beginning of the word whose peak is higher than the peak of the previous word.

This system represents only a starting point for the Serbo-Croatian intonation system analysis and annotation conventions. At this point, the system presented here has not been subjected to testing by other labellers, spontaneous speech, or different dialects. We expect that as we pursue these further tasks, the system will require adjustments and reanalysis.

REFERENCES

Beckman, M. E., and Hirschberg, J. (1993), 'The ToBI Annotation Conventions', ms Ohio State University and AT&T.

Brown, W., and McCawley, J. (1965), 'Srpskohrvatski akcenat', *Zbornik za Filologiju I Lingvistiku*, 8: 147–51.

Godjevac, S. (2000*a*), 'An Autosegmental/metrical Analysis of Serbo-Croatian Intonation', in C. Roberts, J. Muller, and T. Huang (eds.), *Ohio State Working Papers in Linguistics:* Varia, Vol. 54 (Columbus, OH: Ohio State University).

—— (2000*b*), 'Intonation, Word Order, and Focus Projection in Serbo-Croatian', Ph.D. dissertation (Ohio State University).

Grice, M., Ladd, D. R., and Arvaniti, A. (2000), 'On the place of Phrase Accents in Intonational Phonology', *Phonology*, 17/2: 143–85.

Gvozdanović, Jadranka (1980), *Tone and Accent in Standard Serbo-Croatian* (Vienna: Verlag Der Osterreichieshen Akademie Der Wissenschaften).

Inkelas, S., and Zec, D. (1988), 'Serbo-Croatian Pitch Accent: The Interaction of Tone, Stress, and Intonation', *Language*, 64/2: 227–48.

Kostić, Dj. (1983), *Rečenička Melodija u Srpskohrvatskom Jeziku* (Beograd: Rad).

Lehiste, I., and Ivić, P. (1963), 'Accent in Serbo-Croatian: An Experimental Study', in *Michigan Slavic Materials* 4 (Ann Arbor: University of Michigan, Department of Slavic Languages and Literatures).

——, —— (1986), *Word and Sentence Prosody in Serbocroatian* (Cambridge, MA: MIT Press).

Nikolić, B. (1970), *Osnovi Mladje Novoštokavske Akcentuacije* (Beograd: Institut za Srpskohrvatski Jezik).

Pierrehumbert, J. (1980), 'The Phonetics and Phonology of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

——, and Beckman, M. (1988) *Japanese Tone Structure* (Cambridge, MA: MIT Press).

Smiljanić, R., and Hualde, J. (2000), 'Lexical and Pragmatic Functions of Tonal Alignment', in *CSL* 36 *Proceedings* (Chicago: Chicago Linguistic Society), 469–82.

Stevanović, M. (1989), *Savremeni Srpskohrvatski Jezik* (Beograd: Naučna knjiga).

Venditti, J. J. (1995), 'Japanese ToBI Labelling Guidelines', in *OSU Working Papers in Linguistics*, 50: 127–62.

Ward, G., and Hirschberg, J. (1988), 'Intonation and Propositional Attitude: the Pragmatics of L*+HL-H%', in J. Powers and K. de Jong (eds.), *Proceedings of the Fifth Eastern States Conference on Linguistics* (University of Pennsylvania, Philadelphia, PA), 512–22.

Zec, D. (1994), *Sonority Constraints on Prosodic Structure* (New York: Garland Press).

7

The J_ToBI Model of Japanese Intonation

Jennifer J. Venditti

7.1. INTRODUCTION

This chapter presents an overview of Japanese intonational structure and the transcription of this structure using J_ToBI, a variant of the general ToBI tagging scheme developed for Tokyo Japanese. Since the 'Japanese ToBI Labelling Guidelines' (Venditti 1995) were first distributed, J_ToBI has been used in numerous linguistic and computational contexts as a way to represent the intonation patterns of Japanese utterances. This chapter is intended not as a mere rehashing of the 1995 Guidelines, but rather as a comprehensive discussion of the fundamentals of Japanese intonation and the principles underlying the J_ToBI system.

In Section 7.2, we describe the prosodic organization of Japanese and its intonational patterns.[1] We discuss Japanese prosody from a cross-linguistic perspective, highlighting similarities between Japanese and other languages. Section 7.3 then provides an overview of the J_ToBI system. The discussion assumes the reader has some familiarity with intonation description, and with the general ToBI framework. Section 7.4 points out the differences between this new system and its predecessor, the Beckman–Pierrehumbert model presented in *Japanese Tone Structure* (Pierrehumbert and Beckman 1988). Section 7.5 gives an overview of the efforts toward automatization

The author would like to thank Mary Beckman, Sun-Ah Jun, and Kikuo Maekawa for insightful discussions and comments throughout the ongoing development of the Japanese ToBI system.

[1] This discussion and the J_ToBI system itself rely heavily on the model of Japanese tone structure put forth by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988, *inter alia*), which uses a tone-sequence approach to intonation modelling. However, a few important differences between J_ToBI and the Beckman–Pierrehumbert model will be discussed in Section 7.4. This approach is distinct from the superposition-based models of Japanese intonation (e.g. Fujisaki and Sudo 1971; Fujisaki and Hirose 1984; Venditti and van Santen 2000), which will not be discussed here.

of J_ToBI labelling, as well as the degree of labeller agreement using this system, and Section 7.6 lays out future directions for research on Japanese intonation.

7.2. JAPANESE PROSODIC ORGANIZATION AND INTONATION PATTERNS

7.2.1. *Pitch accents*

Japanese is considered a *pitch accent language*, in that the intonational system uses pitch to mark certain syllables in the speech stream. In this way it is similar to languages like English, which also uses pitch accents in its intonational system. However, there are several fundamental differences between the two. First, Japanese and English differ in the level (lexical vs. post-lexical) at which pitch accent comes into play. In Japanese, pitch accent is a lexical property of a word, and thus the presence or absence of an accent on a particular syllable in a Japanese utterance can be predicted simply by knowing what word is being uttered. Take for example the minimal pair shown in Figure 7.1.

Here, the verb /ueru/ in the phrase *uerumono* 'something to plant' is lexically-specified as unaccented, while that in *ue'rumono* 'the ones who are

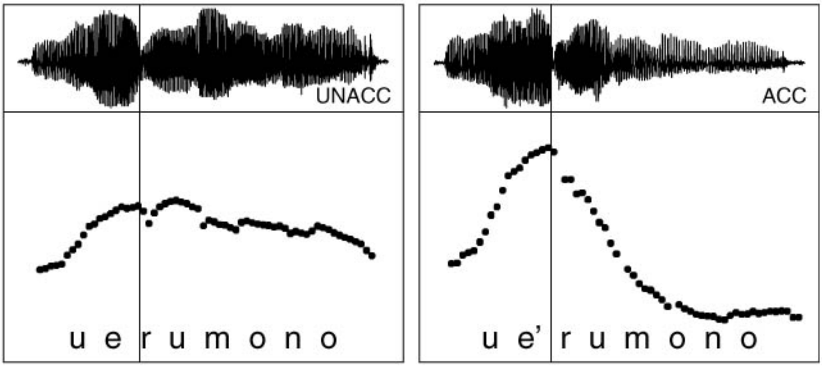


FIGURE 7.1 Waveforms and Fo contours of unaccented *uerumono* 'something to plant' (left) and accented *ue'rumono* 'the ones who are starved' (right) phrases, uttered by the same speaker. The x-axis represents the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contour. Both panels are plotted on the same frequency scale, and vertical lines mark the end of the second mora in each phrase.

starved' is specified as accented on the second mora /e/.[2] The accented phrase displays a precipitous fall in pitch starting near the end of this accented mora, while the unaccented phrase lacks such a fall.[3] This lexical distinction contrasts with languages such as English, in which pitch accents play a role at an entirely different level. In English, the location of metrically strong syllables in a word is determined at the lexical level, and it is these syllables (most often the strongest, or 'primary-stressed' syllable) which serve as docking sites to which pitch accents may be associated at a post-lexical level.

A second difference between the two languages is the function and distribution of pitch accents. In English, pitch accents serve to highlight (or make 'prominent') certain words or syllables in the discourse, and the distribution of pitch accents in an English utterance reflects this function. In a given utterance, there will be a number of metrically strong syllables that can potentially be made even more prominent by the association of a pitch accent. On which of these syllables pitch accents will fall is highly dependent on the linguistic structure of the utterance. That is, an interaction of various factors related to the syntax, semantics, pragmatics, discourse structure, attentional state, etc. will determine where the pitch accents are to be placed in English. In Japanese, in contrast, pitch accents are a lexical property of a given word, and thus they lack any such prominence-lending function. This leaves little room for variability in distribution of accents in a Japanese utterance.

A third difference between the languages is the shapes and meanings of the pitch accents themselves. In Japanese there is only one type of pitch accent: a sharp fall from a high occurring near the end of the accented mora to a low in the following mora. In English, the inventory of pitch accent shapes is far more diverse. There are a number of pitch accent shapes, in which the Fo can rise or fall to/from the accented syllable, or can maintain a local maximum/minimum on that syllable. Each shape has associated with it a specific pragmatic meaning which that accent lends to the overall meaning of the intoned utterance (see e.g. Pierrehumbert and Hirschberg 1990). The Japanese falling accent does not have any such meaning associated with it.

In summary, although both Japanese and English use pitch accent in their intonation system, the languages are in fact quite different with respect to the role that pitch accents play. The languages differ in the level at which pitch

[2] In the transcriptions, accented words contain an apostrophe after the vowel with which the accentual fall is associated; unaccented words lack such a marking.

[3] In the figure, the high to which the Fo rises in the accented case (right) is higher than that in the unaccented case (left). This systematic height difference has been reported in previous studies (e.g. Poser 1984; Pierrehumbert and Beckman 1988; and many others). However, while accented peaks do tend to be higher than unaccented peaks, there is a large amount of variability in both, and there are plenty of cases in read and spontaneous speech where this relative height relation is reversed. Future

accents come into play, in the function and distribution of accents, and in the shapes and meanings of the accents in the inventory.

7.2.2. *Prosodic groupings*

In addition to pitch accents, another important part of Japanese intonation is the grouping of words into prosodic phrases. Speakers can organize their speech into groups of intonational units, which are defined both tonally and by the degree of perceived disjuncture among words within/between groups. This grouping occurs at two levels in Japanese.

First, there is a lower-level grouping, such as that shown in each panel in Figure 7.1. The verb *ueru/ue'ru* is combined with the following unaccented noun *mono* 'thing or person', into a single prosodic phrase. This level of prosodic phrasing in Japanese is termed the *accentual phrase* (AP), and is typically characterized by a rise to a high around the second mora, and subsequent gradual fall to a low at the right edge of the phrase. This delimitative tonal pattern is a marking of the prosodic grouping itself, separate from the contribution of a pitch accent. Both panels in Figure 7.1 consist of a single accentual phrase with the delimitative tonal pattern, though the accented case (right panel) also shows the fall of the lexical accent.[4] The degree of perceived disjuncture between words within an accentual phrase is less than that between sequential words with an accentual phrase boundary intervening. In Tokyo Japanese it is most common for unaccented words to combine with adjacent words to form accentual phrases, though under some circumstances a sequence of accented words may combine, in which case the leftmost accent survives and subsequent accents in the phrase are deleted.

The second type of prosodic grouping in Japanese is the higher-level *intonation phrase* (IP), which consists of a string of one or more accentual phrases. Like accentual phrases, this level of phrasing is also defined both tonally and by the degree of perceived disjuncture within/between the groups. However, the tonal markings and the degree of disjuncture for the IP are different from those of the accentual phrase. The intonation phrase is the prosodic domain within which pitch range is specified, and thus at the start of each new phrase, the speaker chooses a new range which is independent of the former specification. Since there also is a process of *downstep* in Japanese, by which the local pitch height of each accentual phrase is reduced when following a lexically accented

investigations using large amounts of J_ToBI-tagged data are necessary in order to uncover the linguistic factors that are at work in determining this height relationship.

[4] Here, since the accent occurs early in the phrase, the delimitative initial rise is obscured.

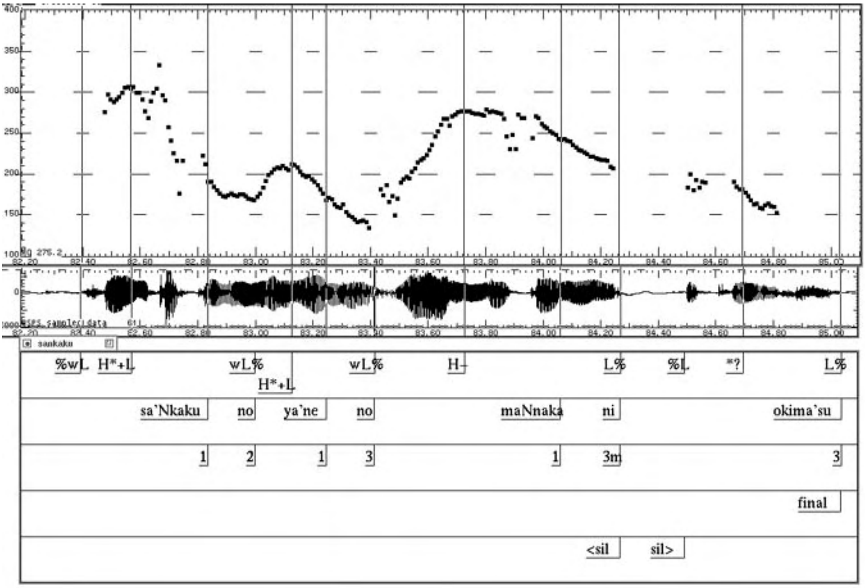


FIGURE 7.2 Fo contour, waveform, and J_ToBI transcription of the utterance ≪sankaku≫: triangle-GEN roof-GEN middle-LOC put 'I will place it right in the centre of the triangle roof'. The x-axis shows the time-course (in sec) of the utterance; the y-axis shows the frequency (in Hz) of the Fo. (Taken from Venditti 1995.)

phrase, one will often observe a staircase-like effect of accentual phrase heights, which is then 'reset' at an intonation phrase boundary. In addition to this behaviour of pitch range, the degree of perceived disjuncture between sequential words across intonation phrase boundaries is larger than that between words within or across accentual phrase boundaries.

Figure 7.2 contains a J_ToBI-transcribed example utterance showing words grouped into accentual phrases and higher-level intonation phrases.[5] The prosodic phrasing of this utterance was judged by a labeller as follows:

accentual phrasing	{ }	{ }	{ }	{ }
intonation phrasing	[	]	[ ]	[ ]
	sa'Nkaku no	*ya'ne no*	*maNnaka ni*	*okima'su*
	triangle-GEN	roof-GEN	middle-LOC	put

[5] At this point, the reader should focus his/her attention only on the Fo contour, the waveform, and the word tier (the 2nd from the top in the label window). A detailed discussion of the symbols in the other label tiers will be presented in following sections.

The accentual phrases *sa'Nkaku no* 'triangular' and *ya'ne no* 'roof-GEN' each are characterized by a rise then rapid fall in the Fo contour. These two APs combine to form the first intonation phrase, with *ya'ne no* being downstepped due to the pitch accent on *sa'Nkaku*, resulting in a staircase-like Fo trend. There is then an expansion of pitch range on the next phrase *maNnaka ni* 'middle-LOC'—this and the virtual pause between *ya'ne no* and *maNnaka* suggest an intonation phrase boundary.[6] The details of the labels in the tone (1st), break (3rd), and other label tiers will be discussed in the following sections.

In addition to the pitch range and disjuncture cues to intonation phrase boundaries, this prosodic unit is also characterized by optional rising or rise-fall tonal movements at its right edge. These movements serve to cue various linguistic and paralinguistic meanings of the utterance, such as questioning, incredulity, explanation, insistence, etc. (e.g. Kawakami 1963/1995; Venditti *et al.* 1998). Each intonation phrase in Figure 7.2 ends in a low tone without such movement, though examples of the various boundary pitch movements occurring in Tokyo Japanese will be discussed in Section 7.3.3.

This section has described the two levels of prosodic grouping in Japanese intonation: the accentual phrase and the intonation phrase. Each of these levels has analogues in other languages as well. Languages as diverse as French and Korean also have tonally-delimited groupings of words like the Japanese accentual phrase (Jun 1993; Jun and Fougeron 1995), and an even larger number of languages have boundary pitch movements which occur at the edge of larger prosodic units analogous to the intonation phrase. Of course, the specific tonal markings used in each language may differ.

English has intonation phrase boundary pitch rises that cue meanings such as questioning and continuation. However, unlike Japanese, English does not have a level of prosodic grouping analogous to the accentual phrase, though the pitch accents of English have a function similar to that of phrasing and pitch range variation in Japanese (see e.g. Venditti *et al.* 1996; Venditti 2000). As mentioned above, since the Japanese pitch accent is hard-coded into the lexical specification of a word, there is little room for variability in pitch accent distribution, as in English. However, the grouping of words into both accentual and intonation phrases (and the pitch range specification of those phrases) is dependent on an interaction of various factors such as the word accentuation, syntactic branching structure, focus, discourse structure, or attentional state, etc.—just those factors affecting English, albeit in a different way.

[6] The phrasing of the remainder of the utterance will be discussed below in Section 7.3.5 when we introduce phrasing/tonal mismatches.

This discussion of Japanese prosodic organization and intonation patterns in comparison with other languages is very important from a cross-linguistic perspective. It shows that the intonational systems of otherwise very diverse languages can be remarkably similar to one another, while maintaining their individual differences. These differences may in fact turn out to be the result of differing means to achieve similar goals. However, only more research on a variety of languages will show how far one can take these cross-linguistic comparisons. In this process, it is essential to be able to use a common framework like the ToBI system to facilitate comparison. With such a tool in hand, we will be much more prepared to start sorting out the similarities and systematic differences among various languages.

7.3. OVERVIEW OF THE JAPANESE ToBI TAGGING SCHEME

The J_ToBI intonation labelling scheme is consistent with the design principles of ToBI systems for English (see Silverman *et al.* 1992; Beckman and Hirschberg 1994; Beckman and Elam 1994) and other languages (this volume). As in other ToBI systems, the transcription consists of the speech and Fo records for the utterance, and a set of symbolic labels. The mandatory labels of a J_ToBI transcription are divided into five separate label tiers in which labels of the same type are marked: tones, words, break indices, finality and miscellaneous.[7] Other optional user-defined tiers can and should be added, as appropriate for the focus of research at each particular site.

The following sections describe the symbolic labels used in the various tiers of a Japanese ToBI transcription.[8] As mentioned in the introduction, J_ToBI for the most part closely follows the theory of Japanese tone structure put forth by Beckman and Pierrehumbert (see Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988, *inter alia*), though a few important differences between J_ToBI and the Beckman–Pierrehumbert model will be highlighted in Section 7.4.

[7] At present, some sites do not use the finality tier. This will be discussed further in Section 7.3.8.

[8] The system described here is identical to that outlined in the 'Japanese ToBI Labelling Guidelines' (Venditti 1995). The reader is referred to this work for more details of the transcription procedure (see also Campbell (1997) for an overview in Japanese). In addition, since the writing of this chapter, an extension of the J_ToBI tagging scheme, dubbed X-JToBI, has been developed by Maekawa and colleagues at the National Language Research Institute (NRLI) in Tokyo, for use in tagging their 'Corpus of Spontaneous Japanese' database (see e.g. Maekawa and Koiso 2000; Maekawa *et al.* 2002). This new scheme introduces additional labels that are necessary to transcribe the spontaneous speech phenomena that they have observed. The reader is referred to future work coming out of NRLI to track the development of this new X-JToBI scheme.

J_ToBI is intended as a tool: the entire purpose of the system is to provide a standard for prosodic labelling of diverse speech data, in order to promote continued research on Japanese intonation. The system is primarily qualitative, in that the symbols employed (and their positioning) reflect the phonological contrasts present in the language. As such, it is useful for those wishing simply to describe the intonational organization of Japanese utterances, for example a psycholinguist needing to describe the prosodic phrasing of his/her experimental stimuli. At the same time, J_ToBI can be a quantitative tool as well. A J_ToBI-labelled database can provide a valuable resource for those wishing to do quantitative modelling of Japanese intonation, for example a computational linguist needing to predict the Fo height relationship between the delimitative high and the accent high within an accentual phrase. Thus, Japanese ToBI is a general-purpose prosodic labelling tool that can be used in many different research contexts.

7.3.1. *Lexical accent tone*

The H*+L composite label placed within the accented mora is used to mark the lexical accent in accented accentual phrases. The H* portion indicates that the high part of the falling tone is associated with the accented mora itself, and the following +L indicates that a low occurs at some fixed point afterwards, usually within the following mora. This H*+L accent label is absent in unaccented words.

Figure 7.2 shows a full J_ToBI transcription of the example utterance ≪sankaku≫. In the tone tier (the 1st from the top in the label window), the H*+L labels on *sa'Nkaku* 'triangle' and *ya'ne* 'roof' mark the lexical accents. The downstep of *ya'ne no* is not explicitly marked (as downstep is in English ToBI), since it is entirely predictable from the lexical accent specification of the preceding phrase.

In many cases, the position of the H*+L label will coincide with the location of the actual Fo maximum (or in the case of a plateau, the start of the precipitous fall), as is the case in Figure 7.2. However, it is not uncommon for the peak to occur after the accented mora, but still be perceived as occurring on the accented mora (e.g. see Sugito 1981; Hata and Hasegawa 1988; Venditti and van Santen 2000). In such cases, two labels are placed: the H*+L is labelled within the accented mora, as usual, and an additional < label is used to mark the actual delayed Fo peak. That is, the H*+L label indicates that an accent is phonologically associated with that particular mora, regardless of whether the Fo peak occurs at that point or not. If necessary, the additional < label pinpoints the actual location of this phonological event in the phonetic record. Careful labelling of the actual F0 event in J_ToBI transcribed databases is essential for research on F0 timing and peak alignment, and on systematic pitch range variation across phrases.

7.3.2. *Accentual phrase tones*

As described in Section 7.2.2, the accentual phrase in Japanese is tonally defined by an initial rise to a high around the second mora of the phrase, then subsequent gradual fall to a low at the right phrase edge. This tonal pattern is shown on the unaccented phrase *uerumono* in Figure 7.1 (left panel), and on the phrase *maNnaka ni* in Figure 7.2. The initial phrasal high tone is marked in J_ToBI by placing a H- label on the second mora of the phrase, while the final low boundary tone is indicated by L% placed at the phrase edge.[9] When the accentual phrase follows a pause (as it does in Figure 7.1), an additional delimitative %L tone is marked at the phrase onset, to provide an anchor from which the F0 rises. Thus, the complete tonal transcription of the APs shown in Figure 7.1 is:

unaccented AP	%L H-		L%
accented AP	%L (H-)	H*+L	L%

Although delimitative tones such as these are found in a variety of intonational systems, the specific tones that each system employs (and which syllables the tones are associated to) will vary across languages. In Japanese there is an additional phenomenon that influences the tonal choice: accentual phrase-initial syllables which are either (i) heavy (i.e. two morae) and sonorant, or (ii) accented, display a rise starting from a higher F0 level than phrases starting with unaccented light (i.e. single mora) syllables. This complex difference in syllable weight affecting the F0 contour is encoded in a J_ToBI transcription by using %wL or wL% boundary tones. The %wL is marked at the beginning of post-pausal phrases, while the wL% is used at the right edge of phrases in cases where the next phrase begins with a heavy syllable or initial accent. Other languages have such language-specific phenomena as well, such as the influence of accentual phrase-initial consonant laryngeal features on the F0 contour in Korean (Jun 1993).

The tonal transcription in Figure 7.2 shows the delimitative accentual phrase tones. The utterance-initial phrase *sa'Nkaku no* is marked with a %wL

[9] Note that the H- phrase tone is labelled on all unaccented phrases, and on accented phrases only where the H- is distinguishable from the high of the lexical accent.

preceding and wL% following, due to the heavy accented initial syllable /sa'N/ and the following accented syllable /ya'/. The phrase *ya'ne no* is also followed by a wL%, due to the heavy syllable /maN/ following. Both phrases *maNnaka ni* and *okima'su* are labelled with L% at their right edge, since they are not followed by such a syllable. The final phrase *okima'su* begins with a %L, since it is post-pausal and starts with an (unaccented) light syllable.

The only phrase in this utterance that is marked with the H- phrase tone is the unaccented *maNnaka ni*, which shows a clear Fo peak around the second mora. Had the peak been delayed (as in the H*+L cases described above), the < would have been used to mark the late Fo event. It is often the case that the peak of the accentual phrase-initial H- rise is delayed to the third mora of the phrase, or even later. At present, it is unclear which factors influence this H- peak placement, though in some cases it appears that information status or speech rate may play a role: the peak is more likely to be delayed or undershot in old information, or in faster rates. This is still a very exciting open research question, which hopefully will be systematically investigated as J_ToBI-labelled databases become increasingly available.

7.3.3. *Intonation phrase tones*

The higher-level intonation phrase in Japanese displays tonal markings as well. As mentioned in Section 7.2.2, rising or rise-fall *boundary pitch movements* ('BPMs') often occur at the right edge of intonation phrases. The H% and HL% boundary tone labels are used to mark these BPMs, respectively.

The HL% is a boundary tone used to mark the rise-fall boundary pitch movement often found in the casual speech of younger speakers. Utterances containing this BPM type are often perceived as sounding 'explanatory' (Venditti *et al.* 1998). The H% boundary tone in the J_ToBI scheme described in the 1995 Guidelines is used for any rising BPM, regardless of Fo height, alignment, or meaning distinctions. However, the nature of H% rises in Japanese can be quite diverse (see e.g. Kawakami 1963/1995). For example, consider the two utterances in Figure 7.3: both are identical in segmental make-up (*/hontô ni na'ra no na no/*), and both consist of 2 APs grouped into one IP, with final BPM rise. As such, they have identical J_ToBI transcriptions (%wL H- wL% H*+L L% H%).

The utterances differ primarily in the height to which the Fo rises at the end of the phrase, and in the time-course of this rise. This difference results in a meaning distinction: the high-rising H% boundary tone (left) cues a question interpretation, while the mid-rising H% (right) cues an insisting

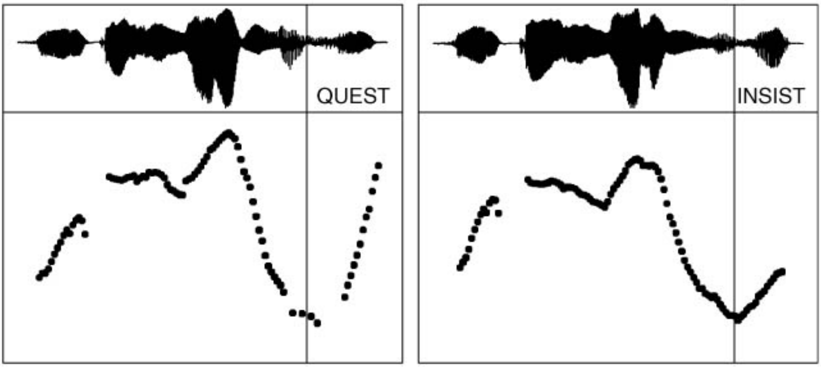


FIGURE 7.3 Waveforms and Fo contours of two productions of *hontô ni na'ra no na no*: really Nara-GEN-COP-QUEST, both uttered by the same speaker with the same tune. The left panel has a question interpretation 'Is it really the one from Nara?', while the right panel has an insisting interpretation 'It's really the one from Nara!'. The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contours. Both contours are plotted on the same Fo scale, and the vertical bars mark the onset of the final mora *no* in each case.

interpretation. Venditti *et al.* (1998) have examined a number of rising BPM types in perception and production studies, and have concluded that the various BPMs in Tokyo Japanese not only cue statistically significant differences in meaning, but can be differentiated by Fo height, rise shape, and timing characteristics as well.

Figure 7.4 shows the shapes of the 5 different BPMs examined in Venditti *et al.* (1998). The figure plots multiple repetitions of raw Fo contours of the phrases *Na'oya ni* 'to Naoya' (left) and *Manami ni* 'to Manami' (right), uttered by a single speaker at a uniform speech rate. The rows show five different BPM types. The contoured lines trace the Fo values of each frame from the start of the phrase to the end of the rise (or the end of the fall in the explanatory type (row 5)). The solid vertical line marks the onset of the final mora *ni* (all contours are time-aligned by this point), and the dashed horizontal line marks a fixed arbitrary Fo reference height. Venditti *et al.* found that rises cueing a question interpretation (rows 1 and 2) are more 'scooped' (concave) and often rise to a higher Fo value than prominence-lending rises or insisting rises (rows 3 and 4). In addition, the timing of rises is different: the rise starts well within the vowel /i/ in *ni* in question BPMs (with the incredulity rise starting latest), while in other BPM types the rise starts at the onset of the final mora of the phrase (*ni*).

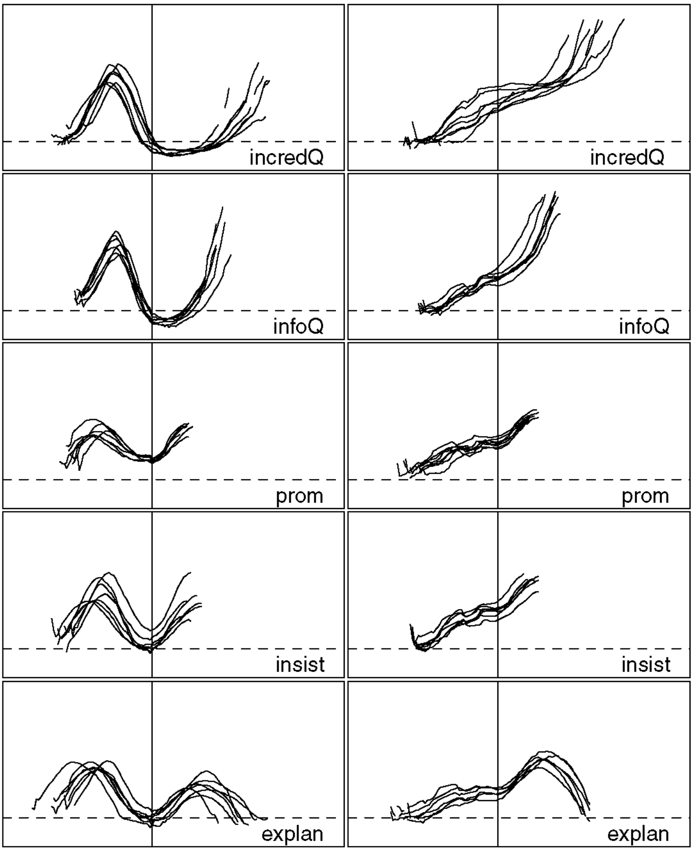


FIGURE 7.4 Fo contours of five boundary pitch movements: incredulity question (row 1), information question (row 2), prominence-lending rise (row 3), insisting rise (row 4), and the explanatory rise-fall movement (row 5). All phrases were uttered by a single speaker at a uniform speech rate on the phrases *Na'oya ni* 'to Naoya' (left) and *Manami ni* 'to Manami' (right). The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contours. All panels are plotted with the same Fo and time scale. (Taken from Venditti *et al.* 1998.)

Under the J_ToBI system described in the 1995 Guidelines, all of the rising utterances (the first four rows) would be transcribed with an H% boundary tone at the right phrase edge. The accented phrase *Na'oya ni* would be transcribed as %wL H*+L L% H%, and the unaccented phrase *Manami ni* would be %L H- L% H%. However, each rise type has been shown to cue a categorically distinct meaning, and the question rises have a different Fo shape than the other two rises; both of these facts suggest that the rises should somehow be distinctly represented in the transcription. Previous studies have shown that differences in pitch range can provide systematic cues to question interpretation in Korean (Jun and Oh 1996) and incredulity vs. uncertainty readings of the L*+H L- H% contour in English (Hirschberg and Ward 1992). In these cases, the phonological tonal transcription is identical in the two interpretations; the only difference is the overall range of the phrase. However, in the case of Japanese BPMs, not only is the pitch range different, but the timing (the alignment of the F0 rise with the segments) is distinct as well. This categorical difference in timing could be encoded in the tonal transcription by introducing an additional LH% boundary tone: in the left (accented) panel of Figure 7.4 both question types show a low region in the final mora preceding the rise (LH%), whereas the prominence-lending and insisting rise types start to rise right at the final mora onset (H%).[10] It is plausible that the low portion of the LH% boundary tone is present in the unaccented question BPMs (right panel) as well, albeit severely undershot. In such a revised system, the new inventory of boundary tones would be as follows:

H%	prominence-lending rise, insisting rise
LH%	incredulity and information question rises
HL%	explanatory rise-fall boundary movement

The difference between rises within each tonal category would then be attributed to differences in pitch range, voice quality, and the like, which do not come into play in a J_ToBI tonal transcription. Increasingly available spontaneous speech databases will be an invaluable resource in order to systematically investigate the acoustic properties of these BPMs, and also to determine their distribution function in connected discourse.

7.3.4. *Marking disjuncture*

Break indices ('BI') are one of the most important parts of a Japanese ToBI transcription, yet for some labellers these may be the most difficult to judge. Break indices are labels indicating the degree of prosodic association between adjacent words or phrases in an utterance. As such, they are primarily subjective values—measures of *perceived* disjuncture between adjacent

[10] The explanatory rise-fall BPM also starts its rise right at the onset of the final mora, which is consistent with the use of the HL% label. In this BPM, there is a marked lengthening of the final vowel (as in questions), which carries both high and low tones.

Table 7.1 Break index levels distinguished by the Japanese ToBI scheme

0	strong cohesion	Typical of fast speech or AP-medial lenition processes (e.g. lenition of a voiced velar stop to an approximant).
1	no higher-level juncture	Typical of the majority of AP-medial word boundaries.
2	medium degree of disjuncture	Typically corresponds to the tonally-defined *accentual phrase* (AP).
3	strong degree of disjuncture	Typically corresponds to the tonally-defined *intonation phrase* (IP).

words—and should therefore be labelled only after careful consideration of the sound record. There are various perceptual cues to disjuncture, including pausing, segmental lengthening, F0 lowering or resetting, creaky voice quality, etc. Listeners certainly can attend to all of these cues when parsing the stream of incoming speech.

The J_ToBI system currently distinguishes four degrees of disjuncture (on a scale from 0 (weak) to 3 (strong)) in the prosodic structure of Japanese.[11] All junctures between words in an utterance are assigned one of these break index values. The levels are summarized in Table 7.1, in order of increasing sense of disjuncture.

Figure 7.2 gives break index labels for the utterance ≪sankaku≫ (see the 3rd tier from the top in the label window). Break index levels 2 and 3 are arguably the most essential, since they show the higher-level prosodic phrasing of the utterance. A medium sense of disjuncture between adjacent words (BI 2) most often corresponds to the tonally-defined accentual phrase boundary. Likewise, a strong sense of disjuncture (BI 3) often corresponds to the tonally-defined intonation phrase boundary. However, there are a fair amount of mismatches between disjuncture and tonally-defined prosodic units, in both read and spontaneous speech. We will discuss these cases in Section 7.3.5.

For the most part, the break index levels and the tonally-defined phrases do match up. This is not a coincidence. As mentioned above, there are many

[11] J_ToBI labelling conducted at ATR in Japan also uses a level 4 break index, which represents an intonation phrase boundary occurring utterance-finally, which has a stronger sense of finality/ completeness than do utterance-medial IP boundaries. However, the system described in the 1995 Guidelines and in this chapter does not include this additional level, but rather delegates this phenomenon to the finality tier (see Section 7.3.8).

perceptual cues to disjuncture, Fo movements being one of them. Unlike lexical accent, phrasing in Japanese allows for some degree of variability, and the prosodic structure that a speaker produces in a given utterance depends on an interaction of a number of linguistic factors, as outlined in Section 7.2.2. One way that speakers cue this prosodic parse (or 'chunking') of an utterance is by tonal movements: words are grouped into accentual phrases characterized by the delimitative tones, and APs are grouped into intonation phrases characterized by a certain pitch range and boundary tones. The initial rise of the accentual phrase cues the start of a new unit, and the pitch range reset at the start of an intonation phrase cues the beginning of an even larger unit. That is, it is the Fo rise itself that provides a major cue to the chunking of an utterance. Therefore, the close relationship between the perceived degree of disjuncture and the tonally-defined prosodic units is not considered circularity in the system, but rather it is a necessary result of Fo rising movements being one of the cues to disjuncture between words.

Another misconception is that labellers' judgements of BI 3 in Japanese is determined solely by the placement of pauses. Although it is true that pausing is often accompanied by the percept of a large degree of disjuncture, this is neither a necessary nor sufficient condition for marking BI 3. For example, there are numerous cases such as that shown in Figure 7.2, in which labellers judge a BI 3 between two words (here, *ya'ne no* and *maNnaka*) where no pause intervenes. As mentioned above, it is likely that the large Fo rise on *maNnaka* (or some other acoustic cues like pre-boundary segmental lengthening, etc.) results in the percept of large disjuncture between the two words. Likewise, there are many cases in spontaneous speech in which a pause is present, but no large disjuncture is perceived. These are cases of hesitations or disfluencies, and are discussed in detail in Section 7.3.6 and Figure 7.6 below.

7.3.5. *Mismatch between tones and perceived juncture*

The previous section described the levels of prosodic association between adjacent words currently recognized in the J_ToBI system. In most cases, break indices 2 and 3 correspond to accentual and intonation phrase boundaries, respectively. However, in some cases there is not such a clear mapping. There are cases in which the perceived degree of disjuncture is appropriate for an accentual phrase break, but there are clear tonal markings of an intonation phrase boundary. Likewise, the degree of disjuncture may seem large, yet the following AP appears to be in a downstepping pattern,

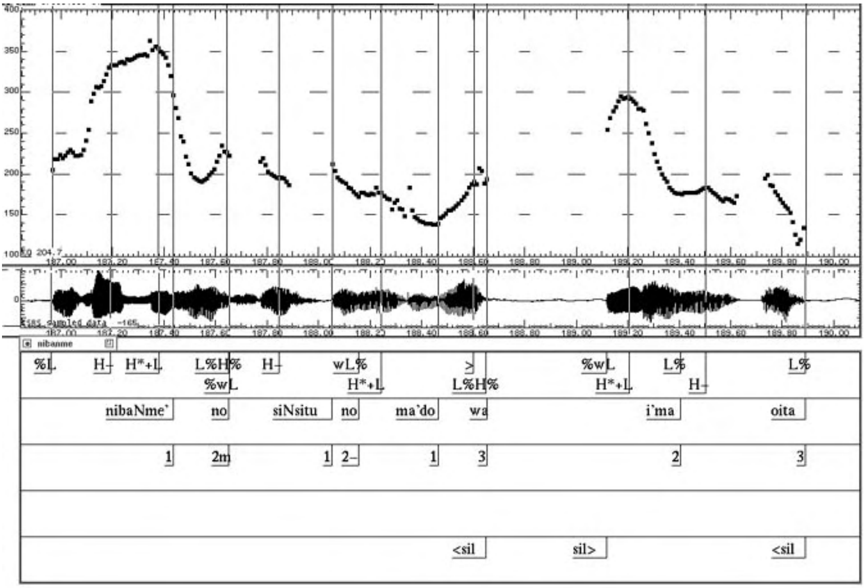


FIGURE 7.5 Sample J_ToBI transcription of the first part of the utterance ≪nibanme≫: second-GEN bedroom-GEN window-TOP now put 'I will put the second bedroom window below the first window which I just laid down'. (Taken from Venditti 1995.)

showing no signs of an intonation phrase break. Figures 7.5 and 7.2 show J_ToBI transcriptions of such cases, respectively.

In Figure 7.5, there is a boundary pitch movement (here, a H% prominence-lending rise) present on the final mora of the first phrase *nibaNme' no* 'second-GEN', suggesting an intonation phrase boundary, but there is no sense of a large disjuncture between this phrase and the following word *siNsitu* 'bedroom'. In fact, the downstepping of *siNsitu* due to the accent in *nibaNme'* suggests that there is no intonation phrase boundary intervening. Figure 7.2 shows another case of mismatch, in which there is a strong break (with pause) after the phrase *maNnaka ni* 'middle-LOC', though the pitch range on the final verb *okima'su* 'put' suggests that there is no intonation phrase break between the phrases. In such cases of mismatch, the break index value is labelled according to the perceived degree of disjuncture, and the accompanying diacritic 'm' is used. Thus, the BI labels in these two examples would be 2m and 3m, respectively.

At present, there are too few data available to conclusively determine what causes such mismatches. In the case of 2m, it is common to observe utterance-medial BPMs in both read and spontaneous speech (e.g. Kawakami 1963/1995; Muranaka and Hara 1994; Nagahara and Iwasaki 1994), especially the prominence-lending rises, and these need not have a pause following or a strong disjuncture. Such a configuration would give rise to a 2m label. As for 3m, this type of contour is often observed in sentence-final position in Tokyo Japanese, in which the verbal predicate is set off from the rest of the sentence by a large juncture preceding, and is produced in a very narrow pitch range.[12] These casual observations about the distribution of mismatches cry out for a more detailed investigation using a large J_ToBI-labelled spontaneous speech database. With such a resource, it will be possible to make better generalizations about when tones and breaks coincide, and when they do not.

7.3.6. *Disfluent junctures*

It is common in spontaneous speech for the speaker to hesitate, stop abruptly and restart, or produce other types of disfluencies. Since the aim of J_ToBI is to describe the intonation of spontaneous as well as read speech, there must be a mechanism for marking such disfluent junctures. Following English ToBI, the diacritic 'p' following a break index value is used to mark these cases. The use of this diacritic on the break index tier is a cue that the corresponding tones on the tone tier may be incomplete or ill-formed.

Figure 7.6 shows three different productions of the fragment *ima no ma'do* 'the livingroom window', uttered by the same speaker in different contexts. The first panel shows a case where there is no disfluency. There are two accentual phrases in sequence: *ima no* 'livingroom-GEN' and *ma'do o* 'window-ACC', with a wL% boundary tone intervening. This internal juncture is label with BI 2. The second and third panels show cases of disfluencies. In both panels, the speaker stops abruptly after the words *ima no*, but then continues on with the following *ma'do* as if no disfluency had occurred (without restart). The difference between the two panels is the strength of the disfluent juncture. In the second panel, there is hardly any sense of disjuncture, and the whole fragment *ima no ma'do to* constitutes a single well-formed accentual phrase in terms of the tones. Thus, the BI value 1 at the disfluency reflects the fact that this juncture falls inside a larger unit (accentual phrase), and the 'p' diacritic flags the disfluency. There is no

[12] Such a contour is strikingly similar in function to the 'finality' contour described in Section 7.3.8, except that it lacks the H% prominence-lending rise. Without the rise, the break is labelled '3m', but with a rise it would be labelled '3'. However, further analyses of more data of this type may show that these are just two variants of the same animal.

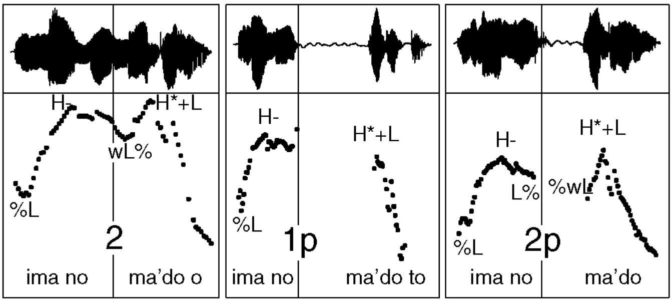


FIGURE 7.6 Waveforms and Fo contours of three productions of the fragment *ima no ma'do* 'the livingroom window', uttered by the same speaker in different contexts. The x-axis shows the time-course of the utterances; the y-axis shows the frequency (in Hz) of the Fo contours. Each contour is plotted on the same Fo scale, and the vertical lines mark the internal juncture. Break indices and tones are labelled for each phrase.

AP-final low tone after *ima no* here. In contrast, the sense of disjuncture in the third panel is stronger, with a clear L% boundary tone realized right before the disfluent region. In this case, the stronger juncture is marked by BI 2, and the 'p' flags the disfluency.

7.3.7. *Labeller uncertainty*

Japanese ToBI allows for marking of labeller uncertainty of both lexical accent realization and break index value. Accent uncertainty is most commonly found in regions of extremely reduced pitch range—for example, cases in which the pitch range of a phrase has been compressed due to the downstepping effect of a preceding accent, and/or by pragmatic or discourse factors. In these cases, the range is so compressed that the lexical accent (cued primarily by the sharp fall in Fo) is hardly perceptible.[13] Such cases are often observed sentence-finally in Tokyo Japanese (see description of the 'finality' contour in Section 7.3.8). Figure 7.7 shows an example of such accent uncertainty. The sentence-final verb *okima'su* 'put' is lexically specified as

[13] But see the production study reported in Maekawa (1994) which shows that words containing 'degenerate accents' (those accents that are realized in a highly reduced pitch range and are often marked with the *? uncertainty label) differ systematically (albeit subtly) from unaccented words in their Fo slope. In addition, Maekawa (1997) presents data which show that such subtle differences in Fo slope can indeed bias listeners' accented vs. unaccented judgements in an identification (perception) task.

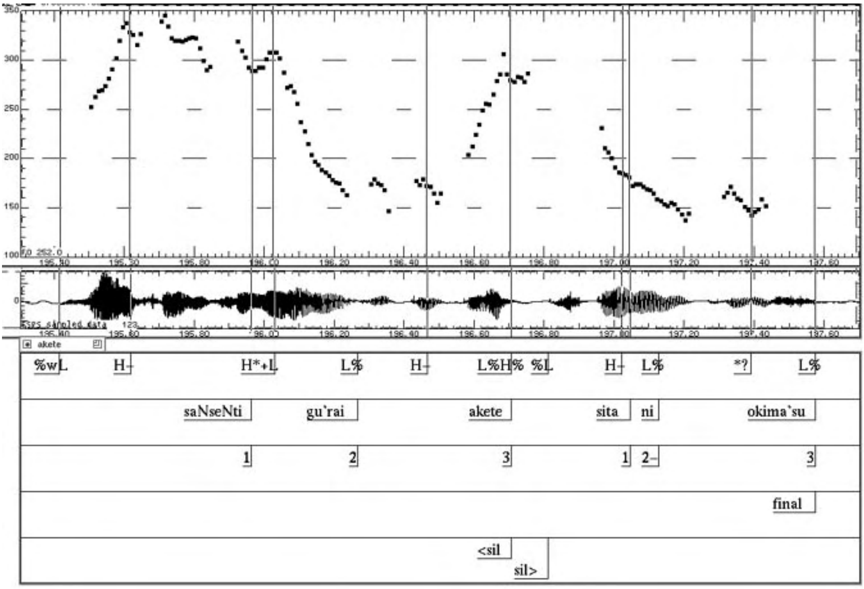


FIGURE 7.7 Sample J_ToBI transcription of the utterance ≪akete≫: 3 cm about open below-LOC put 'I will open up about a 3cm space and put it below there'. (Taken from Venditti 1995.)

accented, but the labeller is uncertain about whether the speaker indeed produced an accent in this case. The '*?' label is used to mark the uncertainty.

In regions of extremely reduced pitch range, not only is the fall of the lexical accent difficult to perceive, but also the signature initial rise of the accentual phrase can be obscured as well. That is, the labeller may find it difficult to judge whether the target word is produced as a separate accentual phrase, or dephrased together with the preceding material to form one single accentual phrase. Such cases lead to break index uncertainty judgements, as shown in Figure 7.7. The labeller is not only uncertain of the accent realization on *okima'su* 'put', but is also uncertain about whether there is an AP break (BI 2) between this and the preceding *sita ni* 'below-LOC'. Break index uncertainty is labelled by adding the diacritic '-' after the break index value, here '2-'.

As with break index judgements themselves, BI uncertainty is highly subjective. Upon careful examination of the sound and Fo records, if the labeller still cannot decide whether or not an accentual or intonation phrase break occurs, the uncertainty label may be used. Uncertainty about whether there is an accentual phrase break (i.e. a medium degree of disjuncture) is labelled by '2-', and uncertainty about larger breaks is labelled by '3-'. That is, the break index value reflects the highest plausible level of phrasing for that particular juncture, and the '-' diacritic marks the uncertainty.

In Japanese ToBI labelling, uncertainty is a *good* thing. If all breaks were easily categorized, the labelling system would not be as meaningful. The uncertainty labels serve as flags to mark areas of interest for future research using large tagged databases, and as such should be used liberally.

7.3.8. *Finality*

The perceived finality of intonation phrases is marked on a separate finality tier. At present this is a simple binary choice between 'final' and 'not final' (no label is used in non-final cases): a phrase which is judged as 'final' will have at its right edge a strong sense of disjuncture, stronger than that of a non-final intonation phrase boundary. The notion of 'finality' is subjective by nature, and will depend on several acoustic and stylistic factors which, in combination, cue that a given phrase is final. These factors include, but are not limited to: final Fo lowering, segmental lengthening, creaky voice, amplitude lowering, long pauses, stylized 'finality' contours, etc.

The utterance ≪akete≫ shown in Figure 7.7 provides an example of finality marking. Here, the last intonation phrase *sita ni okima'su* 'put it below there' is marked with the finality label at its right edge (in the 4th tier from the top in the label window). This utterance is a good example of the so-called stylized 'finality' contour, which is often employed to signal the end of a turn or unit (common in narrative or instructional sequences). In this type of stylized contour, there is typically an H% prominence-lending rise at the edge of the phrase just before the final predicate (note the H% on *akete* here), followed by an optional pause. The final phrase (i.e. the predicate) is realized in a very reduced pitch range.[14] This particular combination of high pitch immediately preceding a very low predicate is often used in Tokyo Japanese to cue the finality of an utterance.

The 'finality' label was introduced into J_ToBI in order to mark turn or unit-final intonation phrases: the tonal pattern of the IP is the same as in other non-final cases, but it somehow has the sense that the speaker is 'done'. This label is found often in sentence-final contexts, but can also be used on medial IPs, especially in extended monologues, where the speaker composes several higher-level units of thought within one 'utterance'. Sites that choose not to include a finality tier in the J_ToBI transcription may mark the finality

[14] In addition to the H% boundary marking, a very prominent accent or unaccented phrase, followed by a predicate with extremely reduced range, can also serve to cue finality.

of intonation phrases by a break index 4 on the break index tier. This is essentially equivalent to a BI 3 marking on the break index tier and 'final' label on the finality tier. However, we recommend that a separate finality tier be used. Although the notion of 'finality' is at this point only vaguely defined, we anticipate that marking in this tier will be modified and further developed by sites whose focus is on the various degrees of finality in discourse planning and production.

7.4. DIFFERENCES FROM *JAPANESE TONE STRUCTURE*

The J_ToBI model of Japanese intonation borrows heavily from the theory of Japanese tone structure put forth by Beckman and Pierrehumbert more than a decade ago (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988). However, there has been a significant amount of research on Japanese intonation since that time, and these new findings, as well as some reanalyses of previous assumptions, have made their way into the current Japanese ToBI model. This section briefly describes the major differences between the two frameworks.

Probably the most noticeable difference between *Japanese Tone Structure* (henceforth 'JTS', Pierrehumbert and Beckman 1988) and J_ToBI is the reduction in the number of prosodic phrase levels. JTS proposed three levels above the word in the prosodic hierarchy of Japanese: the *accentual phrase* (AP), the *intermediate phrase* (iP), and the *utterance* (utt). The accentual phrase was defined exactly as it is in J_ToBI, as a low-level prosodic grouping delimited by the H- and L% tones. While this level of phrasing made it into J_ToBI virtually untouched, the JTS intermediate phrase and utterance have been merged into one level of phrasing in J_ToBI: the *intonation phrase* (IP).[15]

Arguments in JTS for the utterance level were based on the distribution of final H% boundary tones and final lowering: both said to occur utterance-finally. However, most of the data examined in the JTS experiments were short read speech utterances, which lacked the diverse phrasing patterns found in spontaneous speech. It turns out that H% and other boundary pitch movements are extremely common (even most common) utterance-medially, where they appear at the ends of the JTS intermediate phrases (e.g. Kawakami 1963/1995; Nagahara and Iwasaki 1994; Venditti *et al.* 1998). In addition, the utterance-final Fo lowering phenomenon is seen to occur in other

[15] In addition, J_ToBI has borrowed from the English ToBI system the notion of perceived degree of disjuncture (break indices), which also contributes to the definition of AP and IP levels in J_ToBI. This was not present in the JTS framework.

('utterance-medial') contexts as well. In spontaneous speech, there is no clear notion of an 'utterance', and within a given speaker's turn, there may be a number of instances (and degrees) of 'finality' cued by lowering, as mentioned above in Section 7.3.8. Without these two arguments for a separate utterance level, we are left with the JTS intermediate phrase as the highest level of prosodic organization currently motivated for Japanese.

Japanese ToBI has adopted a slightly revised definition of this intermediate phrase. Specifically, in the new system, boundary tones associate to this level of phrasing, and it is the unit marked with the optional 'final' tag in the finality tier. Since this level is no longer 'intermediate' to anything, and in order to emphasize that its definition has been revised, J_ToBI calls this level the *intonation phrase* (IP). This turns out to be a convenient renaming, since the same name is given to high-level prosodic phrases in other languages (e.g. English or Korean), which are also characterized by boundary tones.

Another difference between JTS and J_ToBI is the inventory of boundary pitch movement types. JTS recognized only the H% high-rise used in question contexts, while the 1995 J_ToBI Guidelines added to this by introducing the H% mid-rise in insisting utterances, and the HL% explanatory pitch movement used most frequently by young speakers. In addition, based on the discussion in Section 7.3.3, this inventory can be supplemented even further with the LH% scooped rise. Therefore, three distinct BPMs, H%, LH% and HL%, are currently included in the J_ToBI tonal inventory.

The Japanese ToBI system also introduces a number of labels and diacritics that are necessary to describe spontaneous speech (and which turn out to be useful for read speech as well). The mismatch label 'm' is an extremely important label in the J_ToBI system, as well is the '*?' and '-' labels to show uncertainty. The 'p' label is useful for disfluent breaks, and the various tags on the miscellaneous tier mark regions of disfluencies or other non-speech phenomena.[16] The late and early Fo event labels (< and >, respectively) are also new to the J_ToBI labelling scheme, and are essential for research on Fo timing, alignment, and pitch range variation.

7.5. AUTOMATIZATION AND LABELLER CONSISTENCY

This last section discusses more practical issues in Japanese ToBI labelling: To what extent do labellers actually agree on the J_ToBI transcription of a given

[16] The labels used in the miscellaneous tier are not described in this chapter. The reader is referred to the original 1995 Guidelines (Venditti 1995) for discussion of these, and for more details about the other labels and tiers.

utterance? Can this time-consuming labelling process be automated, even partially? Computer-guided prosodic labelling can potentially be a valuable tool for tagging large databases.

Fortunately, some parts of a J_ToBI transcription can be easily predicted from text. Since many of the tone labels are either lexically-specified, or are delimitative markings which are fixed (phonologically) in both location and type, they are entirely predictable given an accent-coded dictionary entry, as well as a record of the prosodic phrasing of the utterance. These tones include: the lexical accent H*+L, AP-initial H-, AP-final L%/wL%, and the AP-initial (post-pausal) %L/%wL (six of the nineteen J_ToBI labels).

However, the remaining thirteen of nineteen J_ToBI labels are not easily predictable from text alone. Tones which will be difficult to predict include: the intonation phrase boundary tones H%, LH% and HL%, whose location is predictable from phrasing but whose type is dependent on the meaning of the utterance; the early (>) and late (<) Fo event labels, which surely require human-labelling (or a very clever peak-picking algorithm); and the accent uncertainty '*?' label. In addition, even the predictable tone labels crucially assume that the prosodic phrasing of the utterance is known. However, break indices (and their accompanying diacritics) are not entirely predictable from text. As a first attempt, BI prediction could be facilitated by an algorithm which first assigns BI 1 as a 'default' for all junctures, then tries to determine the other BI values in a variety of ways. BI 0 prediction could be facilitated by comparing spectral slices of the uttered speech to categories of slices stored in a codebook for that speaker. BI 2 and 3 prediction could be facilitated by examining the distribution and degree Fo rising movements in the utterance, or by developing a text analysis model given the factors we know to affect phrasing (see Section 7.2.2).

Campbell (1996) describes an attempt at automatically predicting break indices by using a method whereby the phone sequence of the input text is generated, then aligned with the speech signal using text-to-speech and speech recognition tools. The system uses this alignment of the phones (and their durations) and the original Fo contour as input to a text-to-speech intonation module, in order to predict a number of candidate intonation contours and tone/break parses. The candidates are then compared with the original contour to select the optimal J_ToBI parse. Prediction of human-labelled break indices using such a method yielded promising results in Campbell's study: 68 per cent of the junctures were predicted exactly, 69 per cent were matches if the presence or absence of BI diacritics are relaxed, and the agreement rose to 90 per cent if the predicted break indices fell within +/− 1 BI of the human-labelled value.[17] The same study examined human-human break index agreement as well. The labels of two expert labellers were compared for a subset of fifty of the 503 utterances used above (containing 282 junctures), again using only BI levels 2–4. Agreement was very high: 92 per cent of labels were an exact match, while 95 per cent matched when relaxing BI uncertainty. Campbell notes that this high degree of human-human agreement could either be due to the uniform reading style of the sentences, or a break index scale which doesn't allow for individual interpretation of juncture strengths, or both.

Another set of human-human labeller consistency data for break indices is also now available. In addition to the fifteen example transcriptions in the 1995 Guidelines, there are also ten un-transcribed practice utterances included, which labellers can use to get acquainted with the system. These utterances contain a total of 89 junctures, which were labelled by five labellers using BI 0–3.[18] Agreement was calculated across all possible pairs of transcribers for each juncture for each utterance, as has been done in English labeller agreement studies (Silverman *et al.* 1992; Pitrelli *et al.* 1994). The 89 junctures examined here do not include utterance-final junctures. The labeller agreement results are reported in Table 7.2.

Results from two subsets of the data are reported, for three separate definitions of what it is to be a break index 'match'.[19] The first row shows results from all (89) utterance-medial junctures, while the second row is a more limited set of cases (55) in which at least one labeller judged the BI value to be different from '1'. BI 1 could be considered a 'default' value (no sign of a higher-level juncture nor of lenition), and is most commonly marked between a noun and its following postposition. This can potentially be confounded by the definition of a 'word' in Japanese, and so it is not of as much interest in judging labeller agreement of higher-level junctures, which are arguably the ones absolutely essential in the characterization of Japanese

[17] These data are of 503 read utterances J_ToBI-labelled by at least one labeller, and contain 3395 human-labelled junctures. BI 4 labels are included in the tabulation, although these occur utterance-finally, and as such should be totally predictable. It is important to note that this prediction and reported agreement is based on label BI values 2–4 only (excluding BI 0 and 1), so that the high performance in prediction in the +/− 1 BI case is probably a result of most of the data being pooled.

[18] Of the five labellers participating, two were the same expert labellers used in Campbell's (1996) study, one was the author of the Guidelines, and the remaining two had familiarity with Japanese intonation analysis but did not have much hands-on experience with the J_ToBI system. We thank Nick Campbell for providing the time and resources to make this study possible.

[19] We report only on BI agreement here, since a comparison of tonal labels warrants an extended study. Many tones in the J_ToBI system are determined by prosodic phrasing decisions, so it is not useful to compare tonal transcriptions without considering the labeller's prosodic phrase parse of the utterance. A detailed study that takes into account labellers' judgements of breaks, coupled with their tonal markings, is needed. We leave such an analysis of the current data for future work.

Table 7.2 Results of the labeller agreement study

data subset	exact match	relaxing diacritics	within +/− 1
all BI	66%	79%	97%
higher-level BI	46%	67%	94%

prosody. Therefore, the 2nd row in Table 7.2 is considered a more revealing estimate of labeller agreement. The first column reports percentage of exact matches, the second column shows the percentage of matches when relaxing the presence or absence of the BI diacritics '-', 'm' and 'p', and the third column shows the percentage of matches when relaxing these and allowing for agreement within +/− 1 break index value. Although results from this comparison cannot be directly compared to Campbell's results or the results for English ToBI agreement (because of differences in materials, BI inventory, tabulation, etc.), they do show that there still is a fair amount of disagreement among labellers. This could be due to a number of things, such as the complexity of the spontaneous speech testing materials themselves, labeller training, or individual differences in BI interpretation. Hopefully, future studies of labeller agreement, using an increased amount of data and number of labellers, will be able to shed more light on the nature of this disagreement.

7.6. SUMMARY AND FUTURE DIRECTIONS

This chapter has presented an overview of Japanese prosodic structure, and has described the tagging of intonational patterns associated with this structure. We have provided details of the labels used in a Japanese ToBI transcription, along with a discussion of the motivation for, and issues concerning, many of the labels. This system was compared with its predecessor, the Beckman–Pierrehumbert model of Japanese tone structure. Finally, we described efforts toward the automatization of J_ToBI, and summarized results of labeller agreement studies.

It is important to reiterate that Japanese ToBI is first and foremost a research tool, intended to be used to tag intonational patterns in databases of both read and spontaneous speech, in order to facilitate and promote continued research on Japanese prosody. The symbolic labels and annotation conventions currently used in J_ToBI are not etched in stone, but rather are open to improvement and revision, based on new insights gained from the ever-increasing amount of data and analyses available from ongoing research on Japanese intonation. There are many exciting areas of research for which J_ToBI-labelled databases are an invaluable resource. This chapter has mentioned only a handful of such areas: linguistic factors influencing prosodic phrasing, cross-linguistic generalizations, timing and relative height relation of the lexical accent and high phrase tone, boundary pitch movement inventories and their acoustic characteristics, tone/juncture mismatches, stylized (finality) contours, systematic pitch range variation and degrees of finality in discourse, etc. There certainly are many more.

APPENDIX: SUMMARY OF J_ToBI LABELS

H*+L	*Lexical accent*: marked on lexically-accented APs within the accented mora.
<	*Late F0 event*: marked on the actual F0 peak (or start/end of F0 shoulder) when it occurs after H*+L or H-.
*?	*Accent uncertainty*: marked on the lexically-accented mora. Indicates that the labeller is unsure if the accent has been realized.
H-	*AP-initial high phrase tone*: marked on the second mora of the accentual phrase.
L% / wL%	*AP-final low boundary tone*: marked at the right edge of the accentual phrase. The wL% variant is used when the following mora is: (1) heavy and sonorant, or (2) accented.
%L / %wL	*AP-initial low boundary tone*: marked on post-pausal accentual phrases at the leftmost edge. The %wL variant is used when the following mora is: (1) heavy and sonorant, or (2) accented.
H%	*IP-final rise*: marked on the right edge of intonation phrases ending in a prominence-lending or insisting rise.
LH%	*IP-final rise*: marked on the right edge of intonation phrases ending in a question (incredulity or information) rise.
HL%	*IP-final rise-fall*: marked on the right edge of intonation phrases ending in an explanatory rise-fall BPM.
>	*Early F0 event*: marked on the actual F0 peak when it occurs before an H%, LH% or HL%.
0	*Break index*: *strong cohesion*: typical of fast speech or AP-medial lenition processes.
1	*Break index*: *no higher-level boundary*: typical of the majority of AP-medial word boundaries.

2	*Break index*: *medium disjuncture*: typically corresponds to the tonally-defined accentual phrase boundary.
3	*Break index*: *strong disjuncture*: typically corresponds to the tonally-defined intonation phrase boundary.
-	*Break index uncertainty*: marked after the BI value. Indicates that the labeller is unsure of the juncture strength.
p	*Disfluent juncture*: marked after the BI value. Indicates that the juncture is somehow disfluent.
m	*Mismatch*: marked after the BI value. Indicates a mismatch between tones and the degree of disjuncture.

REFERENCES

Beckman, M. E., and Elam, G. A. (1994), 'Guidelines for ToBI Labelling', ms Ohio State University (Version 3.0, March 1997, downloadable from: ling.ohio-state.edu/Phonetics/etobi_homepage.html).

——, and Hirschberg, J. (1994), 'The ToBI Annotation Conventions', ms Ohio State University and AT&T Bell Laboratories.

——, and Pierrehumbert, J. B. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–309.

Campbell, N. (1996), 'Autolabeling Japanese ToBI', in *Proceedings of the International Conference on Spoken Language Processing* (Philadelphia, PA), 2399–402.

—— (1997), 'The ToBI (Tones and Break Indices) System and Its Application to Japanese [in Japanese]', *Journal of the Acoustical Society of Japan*, 53/3: 223–9.

Fujisaki, H., and Hirose, K. (1984), 'Analysis of Voice Fundamental Frequency Contours for Declarative Sentences of Japanese', *Journal of the Acoustical Society of Japan*, 5/4: 233–42.

——, and Sudo, H. (1971), 'Synthesis by Rule of Prosodic Features of Connected Japanese', in *Proceedings of the International Congress on Acoustics*, 133–6.

Hata, K., and Hasegawa, Y. (1988), 'Delayed Pitch Fall Phenomenon in Japanese', in *Proceedings of the Western Conference on Formal Linguistics*, 87–100.

Hirschberg, J., and Ward, G. (1992), 'The Influence of Pitch Range, Duration, Amplitude and Spectral Features on the Interpretation of the Rise-fall-rise Intonation Contour in English', *Journal of Phonetics*, 20/2: 241–51.

Jun, S.-A. (1993), 'The Phonetics and Phonology of Korean Prosody', doctoral dissertation (Ohio State University).

——, and Fougeron, C. (1995), 'The Accentual Phrase and the Prosodic Structure of French', in *Proceedings of the International Congress of Phonetic Sciences* (Stockholm, Sweden), 722–5.

——, and Oh, M. (1996), 'A Prosodic Analysis of Three Types of Wh-phrases in Korean', *Language and Speech*, 39: 37–61.

KAWAKAMI, S. (1963/1995), 'On Phrase-final Rising Tones [in Japanese]', in *A Collection of Papers on Japanese Accent* (Tokyo: Kyûko Shoin Publishers), 274–98.

MAEKAWA, K. (1994), 'Is There "Dephrasing" of the Accentual Phrase in Japanese?', in J. J. Venditti (ed.), *Ohio State University Working Papers in Linguistics*, 44: 146–65.

—— (1997), 'The Intonation of Japanese Interrogatives [in Japanese]', in Onsei Bunpô Kenkyûkai (ed.), *Grammar and Sound* (Tokyo Kuroshio Publishers), 45–53.

——, KIKUCHI, H., IGARASHI, Y., and VENDITTI, J. (2002), 'X-JToBI: an Extended J_ToBI for Spontaneous Speech', *Proceedings of the 7th International Conference on Spoken Language Processing* (Denver, CO: ICSLP), Vol. 3, 1545–8.

——, and KOISO, H. (2000), 'Design of Spontaneous Speech Corpus for Japanese', in *Proceedings of the Science and Technology Agency Priority Program Symposium on Spontaneous Speech: Corpus and Processing Technology* (Tokyo, Japan), 70–7.

MURANAKA, T., and HARA, N. (1994), 'Features of Prominent Particles in Japanese Discourse: Frequency, Functions, and Acoustic Features', in *Proceedings of the International Conference on Spoken Language Processing* (Yokohama, Japan), 395–8.

NAGAHARA, H., and IWASAKI, S. (1994), 'Tail Pitch Movement and the Intermediate Phrase in Japanese', paper presented at the Linguistic Society of America annual meeting, Boston, MA, 6–9 January.

PIERREHUMBERT, J. B., and BECKMAN, M. E. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

——, and HIRSCHBERG, J. (1990), 'The Meaning of Intonation Contours in the Interpretation of Discourse', in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271–311.

PITRELLI, J. F., BECKMAN, M. E., and HIRSCHBERG, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework', in *Proceedings of the International Conference on Spoken Language Processing* (Yokohama, Japan), 123–6.

POSER, W. (1984), 'The Phonetics and Phonology of Tone and Intonation in Japanese', doctoral dissertation (Massachusetts Institute of Technology, Cambridge, MA).

SILVERMAN, K. E. A., BECKMAN, M., PITRELLI, J. F., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J., and HIRSCHBERG, J. (1992), 'ToBI: A Standard for Labeling English Prosody', in *Proceedings of the International Conference on Spoken Language Processing* (Banff, Canada), 867–70.

SUGITO, M. (1981), 'Timing Relationship Between Articulation and Fo Lowering for Word Accent [in Japanese]', *Gengo Kenkyû*, 77.

VENDITTI, J. J. (1995), 'Japanese ToBI Labelling Guidelines', ms Ohio State University. (Also printed in K. Ainsworth-Darnell and M. D'Imperio (eds.) *Ohio State University Working Papers in Linguistics* 50: 127–62 (1997), downloadable from: ling.ohio-state.edu/Phonetics/J_ToBI/jtobi_homepage.html).

—— (2000), 'Discourse Structure and Attentional Salience Effects on Japanese Intonation', doctoral dissertation (Ohio State University).

Venditti, J. J., Jun, S.-A., and Beckman, M. E. (1996), 'Prosodic Cues to Syntactic and Other Linguistic Structures in Japanese, Korean, and English', in J. Morgan and K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Mahwah, NJ: Lawrence Earlbaum Associates), 287–311.

——, Maeda, K., and van Santen, J. P. H. (1998), 'Modeling Japanese Boundary Pitch Movements for Speech Synthesis', in *Proceedings of the 3rd ESCA Workshop on Speech Synthesis* (Jenolan Caves, Australia), 317–22.

——, and van Santen, J. P. H. (2000), 'Japanese Intonation Synthesis using Superposition and Linear Alignment Models', in *Proceedings of the International Conference on Spoken Language Processing* (Beijing, China).

8

Korean Intonational Phonology and Prosodic Transcription

Sun-Ah Jun

8.1. INTRODUCTION

This chapter presents an overview of Korean intonational structure and the most updated version of K-ToBI (Korean Tones and Break Indices) transcription conventions. Korean in this paper refers to Seoul Korean, the standard dialect of Korean. Korean differs from other languages described in this book in that it has neither lexical pitch accent nor lexical stress. Some dialects of Korean such as the Kyungsang dialect have lexical pitch accent as in Tokyo Japanese, but most other dialects including Seoul Korean do not.

Though researchers agree that Seoul Korean does not have lexical stress, it is controversial whether Korean has fixed stress at the word level or phrasal stress. Some believe that Korean has word level stress and that it is sensitive to syllable weight (H.-B. Lee 1964, 1974; H.-Y. Lee 1990; see Lim 2001 for a review); i.e., a word-initial heavy syllable is stressed, and if the first syllable is not heavy, the second syllable is stressed. Thus, according to this view, stress falls on the initial, i.e., first or second, syllable of a word regardless of the word length. Here, 'heavy' is defined as a closed (CVC) or long syllable and is claimed to be acoustically realized with a longer duration.

However, production and perception studies of stress in Korean (Jun 1995*a*; Lim 2001) suggest that the perception of stress on the word-initial syllable is due to the intonation pattern of Korean. Jun (1995*a*) showed that the so-called 'stressed' syllables are always realized with the fundamental frequency (fo) peak when the word is uttered in isolation. When the same word was uttered in utterance-medial position, the 'stressed' syllable of the word showed high fo only when the word was placed at Accentual Phrase-initial position (see Section 8.2.2 for the definition of an Accentual Phrase). The perception test reported in Jun (1995*a*) showed that subjects (seventeen English, two Chinese, one French, one Italian, and one Japanese) perceived the Accentual Phrase (AP) 'initial' syllable as stressed, i.e., prominent. When a word-initial 'stressed' syllable was located in the AP-medial position, the syllable was not produced with high fo, and it was not perceived as prominent. That is, subjects perceived a word-initial syllable as prominent only when the syllable was AP-initial. In addition to the AP-initial syllable, an AP-final syllable was sometimes perceived as prominent.[1] This confirmed the perception of prominence based on high fo because, in Seoul Korean, AP-initial and AP-final syllables are often produced with high fo (see Section 8.2.2 for the phonetic realization of an AP).

Lim (2001) and Lim and de Jong (1999) further illustrate why the perception of prominence or stress has been claimed to be sensitive to syllable weight. They measured the timing of the phrase-initial fo peak when the phrase begins with a heavy or light syllable and found that in general the peak is realized at the end of the first syllable when the syllable is heavy but that it is on the second syllable when the initial syllable is light. This shows that the realization of the fo peak is influenced by the segmental formation of a syllable.

Since these fo peaks correlated with the so-called stressed syllables are due to the Korean intonation pattern, Jun (1995*a*) concluded that the prominence claimed to be a property of a word does not refer to a word level stress but is linked to a phrasal phenomenon, i.e., a by-product of a phrase level prosody.

These different views of stress in Korean have been reflected in the study of rhythm and intonation. The next section describes the intonation studies based on word level stress and phrase level stress. Stress, though hard to define acoustically, was used to define a rhythmic unit in Korean (called 'maltomak' in H.-B. Lee (1964, 1974)), and the view of phrasal level prominence was taken in the intonation phonology of Korean and the transcription of Korean prosody (Jun 1996, 2000) by analysing the phrase-initial fo rise as a phrasal tone, instead of a pitch accent as in English.

The organization of this chapter is as follows. Section 8.2 describes the previous research and intonational phonology of Seoul Korean; Section 8.3 describes the transcription conventions of Korean ToBI (Jun 2000), and Section 8.4 reports the results of labeller agreement and consistency in the transcription of Korean prosody using Korean ToBI.

[1] This supports the observation of Polivanov (1936) and Trubetzkoy (1939) that Korean tends to emphasize a word final syllable, demarcating a word boundary (cited in Koo 1986). This is so because an AP-final syllable coincides with a word-final syllable, and in general one word forms one AP unless the accentual phrasing is influenced by focus, semantic relation, and speech rate, in which case more than one word can form one AP (see Jun 1993, chapter 5 for factors affecting the phrasing; Schafer and Jun 2002 for default phrasing).

8.2. INTONATION OF SEOUL KOREAN

8.2.1. *Background*

Korean is an intonation language. The pitch modulation over an utterance is not specific to a certain syllable of a word, but is a property of a sentence. The intonational contour of a sentence changes the sentence type and the meaning or the information structure of a sentence. For example, a sentence can be interpreted as a declarative if it ends in a Low tone but as an interrogative if ending in a High tone. In addition, the same sentence can be interpreted as a yes-no question or a wh-question depending on the intonational phrasing of the sentence (Jun and Oh 1996). Unlike English or German where f0 peaks and valleys, e.g. pitch accents, are in general linked to the stressed syllable of a word, the peaks and valleys of Korean intonation do not link to any specific syllable of a word but to a certain location of a phrase.

Early studies of Korean intonation (e.g. Martin 1954; H.-B. Lee 1964, 1974; S.-B. Cho 1967) focused on the tonal contour occurring at the end of an utterance, influenced by the tradition of British intonation models. Relying exclusively on auditory impressions, they proposed multiple tonal categories. Martin (1954) proposed seven intonation morphs such as Period intonation, Comma intonation, and Question-mark intonation, etc., while H.-B. Lee (1964) proposed three static (perceptually level pitch) tones and seventeen kinetic (gliding pitch) tones (nine uni-directional tones, four bi-directional tones, and four tri-directional tones). On the other hand, S.-B. Cho (1967) proposed twelve directional intonational forms (four uni-directional, four bi-directional, and four tri-directional forms) with three levels of voice range.

Among these, H.-B. Lee (1964, 1974) was also concerned with a rhythmic unit within an utterance. He proposed 'maltomak' (a rhythmic unit, literally meaning 'a unit of speech') which includes one stressed syllable, optionally preceded and followed by one or more unstressed syllables. This unit can be larger than a word or smaller than a word and is influenced by speech rate and speech style. It is preceded and followed by a pause, large or small. H.-Y. Lee (1990) extended H.-B. Lee's (1974) model by proposing a Rhythmic Group, a rhythmic unit higher than the maltomak. His Rhythmic Group corresponds to an intonation group as well as a breath group; thus, it could be larger than Jun's Intonation Phrase (see Section 8.2.2). However, like H.-B. Lee (1974), H.-Y. Lee's analysis of Korean intonation, which is also based on the impressionistic descriptions adopting the British intonation model (i.e., O'Connor and Arnold 1973), does not provide objective criteria of each prosodic unit. As noted in Seong (1995), the rhythmic unit, maltomak, does not have clear phonetic cues, and some of the domains could only be perceived by a trained phonetician. The abstract nature of maltomak in these studies is partly due to the ambiguous nature of stress in Korean (see Section 8.1) and also partly due to the subjective criteria based on the author's auditory impression in defining the unit.

Koo (1986) is, to my knowledge, the first acoustic study of intonational structure of Korean based on pitch track analysis. He identifies five different patterns from monosyllabic utterances: (1) rise, (2) rise-fall-rise, (3) rise-fall, (4) level, and (5) fall; and three terminal tonal variations from various sentences: (1) rise-fall, (2) large rise-large fall, and (3) rise. Koo also identifies a basic tonal pattern of a small phrase within an utterance, called a 'minor phrase'. His minor phrase is marked by a phrase final rising; thus, it seems to correspond to the Accentual Phrase in Jun's (1993, 1998) model described below.

These earlier studies of Korean intonation, however, described intonation phonetically. They did not assume that fo contours are composed of a sequence of categorically distinct tones which defines a hierarchical prosodic structure at a phonological level. They also did not distinguish intonation, i.e., *linguistic* features of fo, from paralinguistic features of speech (see Ladd 1996, chapter 1, for the definition of intonational phonology). They proposed certain tonal categories based on a speaker's emotional state and his or her attitude towards a hearer. Except for Koo (1986), their analyses were based on auditory impressions and did not have objective criteria of defining an intonational or rhythmic unit.

8.2.2. *Intonational phonology of Korean*

A phonological model of Korean Intonation was proposed by Jun (1993, 1998), based on previous work by de Jong (1989), S.-H. Lee (1989), and Jun (1990). According to this model, which adopts the autosegmental-metrical model of intonation developed by Pierrehumbert and her colleagues (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; see Ladd 1996 for extensive review), the intonational structure of Korean is hierarchically organized in such a way that an Intonation Phrase (= IP) can have more than one Accentual Phrase (= AP), which in turn can have more than one phonological word (= w; a lexical item followed by case markers or postpositions). The AP in Korean is, thus, similar to the Accentual Phrase in Tokyo Japanese (Pierrehumbert and Beckman 1988; Venditti 1995, this volume Ch. 7) or the Accentual Phrase in French (Jun and Fougeron

1995, 1997, 2000, 2002; Fougeron and Jun 1998). It is a tonally demarcated unit which can contain more than one lexical item.

Existing data on intonational phrasing (e.g. Jun 1989, 1990, 1993) suggest that the prosodic units in Korean are hierarchically organized following the Strict Layer Hypothesis (Selkirk 1984; Nespor and Vogel 1986; Hayes 1989). That is, an IP is exhaustively parsed into a sequence of APs, and an IP boundary coincides with an AP boundary which again coincides with a word boundary. The intonational structure of Seoul Korean is schematically represented in Figure 8.1. Categories in the parentheses are optional.

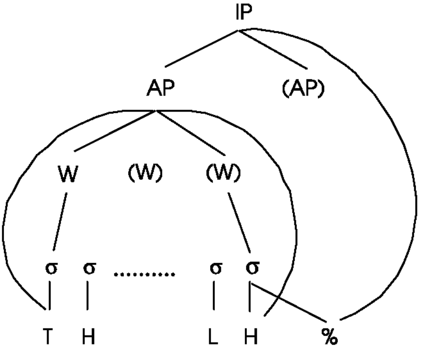


Figure 8.1 Intonational structure of Seoul Korean.

IP: Intonation Phrase AP: Accentual Phrase
w: phonological word σ: syllable
T = H, when the syllable-initial segment is aspirated/tense; otherwise, T = L
%: Intonation Phrase boundary tone

The IP and the AP are two prosodic units in Korean marked by intonation. An IP contour includes tonal patterns of one or more APs and an IP boundary tone. An IP-final syllable is substantially lengthened—about 1.8 times longer than a non-IP-final syllable (Korea Telecom Research and Development Group Report 1996) and optionally followed by pause. An AP has a tonal pattern demarcating the beginning and the end of the phrase. When an AP is final to an IP, the IP-final syllable is realized with the IP boundary tone by preempting the AP-final tone. The first segment of the AP is slightly, and consistently, longer than the same segment in the AP medial position (Jun 1993, 1995*b*; T. Cho and Keating 2001; Keating *et al.* 2004), but the AP final segment is not always longer than the AP medial segment (Koo 1986; Jun 1993, 1995*b*, 1996; T. Cho and Keating 2001). Furthermore, an AP is never followed by a pause unless it is the last phrase of an IP.

A phonological word (= w) has no tonal pattern specific to this level, but it has been shown that, like an AP and an IP, a phonological word in Korean serves as the domain of phonetic strengthening and weakening: VOT (Voice Onset Time) is longer and the linguopalatal contact area of stops is larger at the beginning of each prosodic unit (w, AP, IP) than in the middle of each unit, and VOT is longer at the beginning of a higher prosodic unit than at the beginning of a lower prosodic unit (Jun 1993; T. Cho and Keating 2001; Keating *et al.* 2004).

Finally, like the prosodic units in Greek (Arvaniti and Baltazani this volume Ch. 4), Korean prosodic units defined by intonation are also domains of segmental phonological rules (Jun 1993, 1998). For example, an AP in Korean is the domain of Lenis Obstruent Tensing: a lenis obstruent becomes tense after a lenis obstruent if both are in the same AP but not if there is an AP boundary between the two lenis obstruents. An IP also serves as the domain of phonological rules such as Obstruent Nasalization and Spirantization (Jun 1993).

(i) *The accentual phrase*: the most common tonal pattern of the AP is Low-High-Low-High (LHLH) or High-High-Low-High (HHLH), thus, THLH in Figure 8.1 with T = H or L. The AP-initial tone is determined by the laryngeal feature of the phrase-initial segment: when the segment is either aspirated or tense, having [+stiff vocal cords] (Halle and Stevens 1971), the AP begins with a H tone; otherwise, with an L tone. (For quantitative data about this tonal difference, see Jun 1996; H.-J. Lee and Kim 1997; H.-J. Lee 1997.) When an AP has more than three syllables, the two initial tones of an AP are associated with the two initial syllables of the AP, and the two final tones of an AP are associated with the two final syllables of the AP. The syllables between the second and the penult of the AP, if there are any, get their surface pitch values by interpolating between the H tone on the second syllable and the L tone on the penult. The slope of this falling fo is negatively correlated with the number of syllables within an AP (J.-J. Kim *et al.* 1997). This suggests that the two rises in sequence, AP-initial rise and AP-final rise, are not independent, but belong to the same tonal unit.

The second AP tone, H, is in general realized on the second syllable of an AP (Koo 1986; S.-H. Lee 1989; Jun 1990), but as reported in H.-J. Lee and H.-S. Kim (1997), for speakers who tend to undershoot the penult L tone, the H is sometimes (27–40 per cent) realized on the third syllable in a five-syllable-AP. More recently, however, de Jong (2000) found that the H tends to be realized at the end of the first syllable of an AP when the syllable is a closed syllable. Thus, in this paper, I will adopt Jun's (1993) claim that the second AP tone H is *loosely* associated with the second syllable of an AP.

When an AP has fewer than four syllables, it does not show two rising patterns. Instead, when an AP begins with a non-aspirated/tense segment, it shows a simple rising pattern (LH) or a delayed rising pattern (LLH) or an early rising pattern (LHH) (or for an AP beginning with an aspirated/ tense segments, HH, HLH, and HHH patterns, respectively). Thus, it is assumed in Jun (1993, 1998) that the underlying tone pattern of an AP is THLH and one or both of the two middle tones (i.e., T**HL**H) are undershot when there is not enough time to reach the tonal target. But it is not clear what the conditions for undershooting one of the two middle tones are (e.g. LHH vs. LLH). Observation of data shows that the choice of tones undershot varies across speakers and across different discourse contexts (H.-J. Lee and H.-S. Kim 1997). More importantly, different tonal realizations do not seem to have contrastive meaning.

In addition, the final tone of an AP is sometimes realized as an L tone due to a constraint on the tonal sequence or stylistic variations. This happens whether an AP is long or short. This means that there are at least fourteen surface tonal patterns for an AP (see Figure 8.6 for schematic fo contours of the AP tonal patterns).

(ii) *The intonation phrase*: an IP boundary tone is realized on the IP-final syllable, indicating the pragmatic meaning of the phrase as well as information about the sentence type (Park 2003). Depending on the shape of fo contour starting from the onset of the IP-final syllable, at least nine boundary tones have been identified (L%, H%, LH%, HL%, LHL%, HLH%, HLHL%, LHLH%, LHLHL%). H% and LH% differ in the timing of rising (see Figures 8.8, 8.9, 8.10); LH% rises later than H%, showing a fo valley at the beginning of the IP-final syllable. The same is true with HL% vs. LHL% and HLH% vs. LHLH%. In general, tones ending with H% often have the function of seeking information (e.g. question) and those ending with L% often have the function of making a statement. However, the relationship between a tone and the meaning is many-to-many. That is, more than one boundary tone can be used to mark the same meaning, and the same boundary tone can be used for more than one meaning (e.g. H.-Y. Lee 1997; Park 2003). For example, a wh-question can be marked by L%, H%, LH%, HL%, or HLH% (see Jun and Oh 1996), but each boundary tone can also mark different sentence types or pragmatic meanings. Furthermore, boundary tones delivering the same pragmatic meaning or marking the same sentence type can be different depending on sentence endings. Park (2003) reports that the sentence ending in <-guna> takes HL% while <-ne> takes LH% even though both of these deliver the same meaning of discovery and confirmation (e.g. <zaR bwaD-guna> vs. <zaR bwaD-ne> 'You did a good job!'). More research is needed to identify a distinctive pragmatic meaning for each boundary tone and sentence ending.

8.3. KOREAN-ToBI (K-ToBI)

K-ToBI is a prosodic transcription convention for standard (Seoul) Korean. Like the other ToBI systems, K-ToBI assumes intonational phonology with a close relationship to a hierarchical model of prosodic constituents. The intonational analysis and attendant prosodic model of Seoul Korean adopted for K-ToBI are based on Jun (1993, 1996, 1998). A first version of K-ToBI was developed at ATR Interpreting Telecommunication Systems in Japan in late 1994 by Mary Beckman and Sun-Ah Jun, as part of a Korean synthesis development project. The second version (Beckman and Jun 1996) was proposed at the Prosody Transcription Workshop held just before ICPhS (International Congress on Phonetic Sciences) in Stockholm, August 1995. The current version is a revision of the second version by the author after the Korean ToBI Workshop in Korea, August 1998, and was presented at the workshop 'Intonation: Models and ToBI Labelling', a satellite meeting of ICPhS in San Francisco in August 1999.

The earlier versions of Korean ToBI had four parallel tiers as in the original ToBI system (i.e., Mainstream American English ToBI): words, tones, break-indices, and miscellaneous. But, in order to describe surface tonal patterns which are not always the same as the underlying tonal patterns, and not predictable from the underlying tones, the current version of K-ToBI expands the tones tier into two tiers, a phonological tone tier and a phonetic tone tier. (See the next section for motivation for this change.) Therefore, a K-ToBI transcription for an utterance now minimally consists of a recording of the speech, an associated record of the fundamental frequency contour, and the transcription-proper symbolic labels for events on five parallel tiers (a word tier, a phonological tone tier, a phonetic tone tier, a break-index tier, and a miscellaneous tier).

8.3.1. *Motivation of revision*

In the earlier version of K-ToBI, there were only two tones transcribing the tonal pattern of an AP, i.e, H- and LHa. The H- was labelled to cover any high peaks occurring at the 'initial' (the first or second, and rarely the third) syllables of an AP, and the LHa was labelled to mark the AP boundary ('a' for an AP boundary), which was typically realized as a rising pitch. This transcription, however, turned out to be too abstract and at the same time not distinctive enough.

Labelling LHa at the end of an AP was too abstract when an AP ended in an infrequent low pitch, or when the AP final two syllables do not show a rising pitch (i.e., LHH pattern). On the other hand, H- tone was not distinctive enough. The ToBI labelling system assumes that tones are labelled only when they are distinctive (Beckman and Ayers 1994; Beckman and Hirschberg 1994). However, the realization of AP initial peak is optional, constrained by the length of a phrase and the laryngeal feature of the AP initial segment, and its presence or absence does not seem to change the meaning of the utterance. What is distinctive is the presence or absence of an AP boundary. For example, wh-questions and yes/no-questions are distinguished only by an AP boundary between the wh-word and the following verb phrase (Jun and Oh 1996) and syntactically ambiguous sentences are disambiguated by differences in AP boundary locations (Schafer and Jun 2002). Thus, changing accentual phrasing can change the meaning of a sentence, but whether the AP is realized with initial rise or not does not seem to change the meaning of a sentence.

Furthermore, H-, being labelled at the first occurrence of a high-pitched syllable in an AP did not reflect the different phonetic realization of the peak depending on the origin of the H tone or the alignment of the peak to syllables. The AP initial peak is realized on the first syllable when the syllable begins with an aspirated or tense consonant. In this case, the following syllable is also realized with H tone. The AP initial peak can also be realized later than the first syllable when the first syllable does not begin with an aspirated or tense consonant and when the AP is longer than three syllables. Quantitative data show that fo is significantly higher for the H tone on the first syllable of an AP (i.e., HHLH) than the H tone after AP-initial L tone (i.e., LHLH). (See Figures 8.2 and 8.4 to compare the fo difference between these two H tones.) In addition, this extra-high fo value in the beginning of the HHLH pattern influences the following syllables, if there are any, by raising the fo values of these syllables, compared to those in the LHLH pattern, up to the penultimate syllable of an AP (see H.-J. Lee 1999 for more detail). This suggests that the AP initial peak should be labelled separately for the first and the second syllables, and that these two H tones should be treated differently from the distinctive tones marking the AP boundary.

Therefore, in the current version, we will split the tone tier and label the AP boundaries at a phonological tone tier, and the realization of AP tones at a phonetic tone tier aligned with the corresponding surface fo event. Labelling the surface tones at a phonetic tone tier would also allow us to transcribe the fourteen different surface tonal patterns of AP including the AP final low tone and early rise and late rise (i.e., LH, LHH, LLH, LHLH, HH, HLH, HHLH, LL, HL, LHL, HHL, HLL, LHLL, HHLL). These surface tonal patterns, though seemingly not distinctive, are not fully predictable, and the detailed conditions on the surface patterns or their pragmatic meanings are not yet known. By labelling surface tonal events on a phonetic tone tier, we hope to get answers to these issues and get information about the timing and magnitude of the fo realization of the surface tones.

It should be noted that labelling in the phonetic tone tier should not be interpreted as labelling the gradient phonetic detail of fo contour as in the narrow phonetic transcription such as INTSINT (Hirst and Di Cristo 1998; see Chs. 1 and 2 this volume for the distinction). The tonal inventories in the phonetic tone tier are categorical (i.e., H and L) and their distributions are limited, constrained by the alignment of the AP initial and final two syllables. The data from the phonetic tone tier will provide valuable information of phonetic implementations to researchers working on speech synthesis and recognition and provide feedback about the model to those working on a phonological model of Korean intonation.

Finally, by separating the tone tier into phonological and phonetic tone tiers, we can easily accommodate tonal transcriptions of other dialects. For example, the tonal pattern of an AP in the Chonnam dialect (a southwestern dialect of Korean) is the same as that of the Seoul dialect except that its AP final is falling (i.e., LHL or HHL; Jun, 1989, 1993, 1996, 1998). Though the tonal patterns of APs differ between the two dialects, the accentual phrasing is the same for these dialects. Thus, the boundaries marked on the phonological tone tier of Seoul Korean will remain the same for the Chonnam dialect, while the phonetic tone tier of these two dialects will differ, conforming to the surface realization of each dialect. I assume this will be true for other dialects of Korean which do not have lexical pitch accent.

In the following sections, each of the five tiers is defined, and the labels and symbols proper for each tier are introduced. In addition, example sentences illustrate how to label information on each tier aligned with pitch tracks using *PitchWorks* (Scicon R&D), speech analysis and labelling software similar to *xwaves.* A summary of tones and break indices is given in Appendix A.

8.3.2. *Tiers*

(i) *The words tier*: the words tier in K-ToBI corresponds to the 'orthographic tier' in English ToBI. In this tier, words may be labelled using either Hangul orthography or some conventional Romanization. In the current K-ToBI, words are transcribed following the Romanization convention shown in Appendix B. What constitutes a 'word' in Korean is controversial. In this version, we consider it as a sequence of segments divided by a space in a written Hangul text. The word label should be placed at the end of the final segment in the word, as determined by the waveform or spectrogram record. Filled pauses and the like should also be labelled on this tier.

(ii) *The phonological tone tier*: the phonological tone tier includes the boundary tone of an AP and an IP. Since an AP boundary tone in an IP-final position is overridden by the IP-final boundary tone, only the IP-final boundary tone (%) will be labelled at the end of an IP.

The boundary of an IP-medial AP will be labelled by 'LHa' reflecting the most common AP-final rising tone in Seoul Korean. An IP final boundary will be labelled by one of the nine different boundary tones: H%, L%, HL%, LH%, HLH%, LHL%, HLHL%, LHLH%, LHLHL%. Instructions on where to put phonological tone labels are given below. To simplify the description of IP boundary tones, 'T' is used below as a variable for the IP boundary tones. The meaning of each boundary tone and example sentences labelled with phonological tones are given in the next section.

LHa marks the end of an IP-medial AP, aligned with the end of the AP-final segment determined from the waveform.

T% marks the end of an IP, aligned with the end of the IP-final segment determined from the waveform. 'T' can be H, L, HL, LH, HLH, LHL, HLHL, LHLH, or LHLHL.

(iii) *The phonetic tone tier*: the phonetic tone tier includes the surface tone patterns of APs and IPs. For IP tones, there are nine boundary tones. For AP tones, there are three initial tones (i.e., L, H, and +H) and three final tones (i.e., La, Ha, and L+).

AP-initial tones:

L This tone marks an L tone on the first syllable of an AP. This label should be aligned with the fo valley on the first syllable of an AP.

H This tone marks an H tone on the first syllable of an AP. This label should be aligned with the fo peak on the first syllable of an AP.

+H This tone marks the H tone on the second syllable of an AP (or sometimes the third syllable when the AP is long, uttered quickly, or produced under focus). This label should be aligned with the fo peak around the second syllable. When the peak continues over the following syllable, align this label with the latest fo peak of the phrase-initial peak. This tone is not labelled if both the preceding and the following syllables have a H tone.

Figure 8.2 shows an example pitch track illustrating how to label AP-initial tones on the phonological and phonetic tone tiers. The phonological tone tier is named ‘Utones’; the phonetic tone tier ‘Stones’; and the word tier ‘words’.

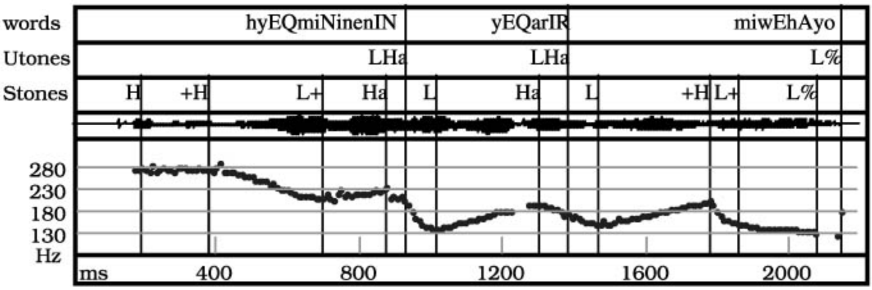


Figure 8.2 An example utterance, hyEQmiNinenIN ‘Hyungmin’s family-TOP’+ yEQarIR ‘Younga-ACC’+miwEhAyo ‘hate’ => ‘Hyungmin’s family hates Younga’, illustrating how to label AP-initial tones. The first AP begins with an H tone, and the second and the third APs begin with an L tone. +H is shown in the first and the last APs.

AP-final tones:

Ha This tone marks either the end of a rising tone or a high flat tone. It is the most common AP-final tone. This label is aligned with an actual f0 peak on the AP-final syllable.

La This tone marks either the end of a falling tone or a low flat tone. This AP-final tone is less common. This label is aligned with an actual f0 valley on the AP-final syllable.

L+ This tone marks the low pitch on the penultimate syllable of an AP. This tone is not labelled if the low pitch on the penult is predictable from adjacent tone labels (e.g. when an AP is continuously falling from an initial H to a final La or when an AP-initial tone is L and the final tone is La). When not predictable, this label is aligned with an actual f0 valley on the penult of an AP. When there is a low plateau adjacent to the penult, place this label at the transition point, i.e., H to L or L to H (see Figure 8.4).

The ‘+’ sign in Korean ToBI, i.e., +H and L+, refers to a syllable boundary and implies a grouping of tones: +H is part of the AP-initial tone realized on the second syllable of an AP and L+ is part of the AP-final tone realized on the penult of an AP. This is different from the ‘+’ in English bitonal pitch accents such as L+H* or L*+H, where both tones are associated with a stressed syllable with the unstarred tone being realized either before the starred tone (i.e., a leading L tone in L+H*), or after it (i.e., a trailing H tone in L*+H). Figure 8.3 shows an example of AP-final tones, Ha, La, L+. Figure 8.4 shows examples of L+ before an AP-final rise and an L+ before L%.

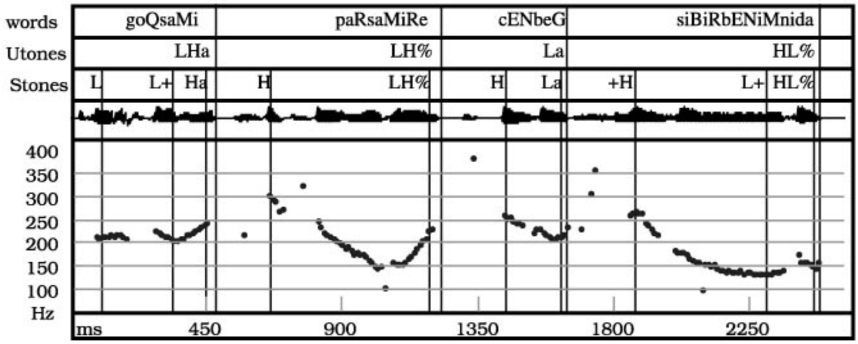


Figure 8.3 An example utterance, goQsaMi '032'+paRsaMiRe '831-and'+cENbeG siBiRbENiMnida '1111-number-is' => '(The phone number is) 032-831-1111', illustrating AP-final tones: Ha (1st AP), La (3rd AP), and L+ (1st and last AP).

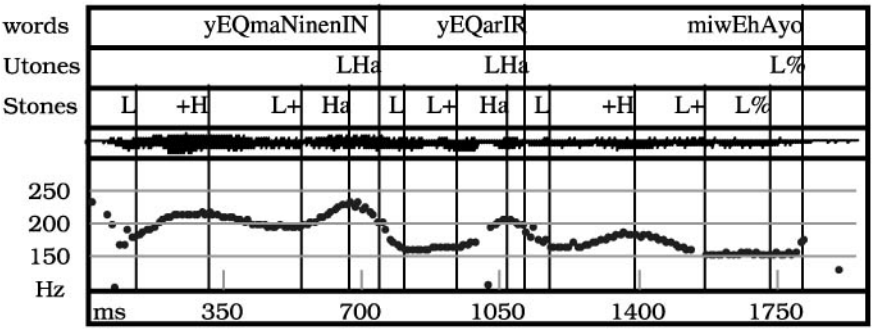


Figure 8.4 An example utterance, yEQmaNinenIN 'Youngman's family-top'+ yEQarIR 'Younga-acc'+miwEhAyo 'hate' => 'Youngman's family hates Younga', illustrating how to label an AP-final L+ tone.

When an AP has three tones, the mid tone can be either L (e.g. LLH) or H (e.g. LHH). In this case, we will consider the medial L as a part of the *final* AP tone and the medial H as a part of the *initial* AP tone because we believe that both are the undershoot version of the underlying two rises, LH-LH. That is, LLH is parsed as L-LH with the undershoot of the first H of LHLH, and LHH is parsed as LH-H with the undershoot of the second L of LHLH. Therefore, LLH will be labelled as L, L+, and Ha, and LHH will be labelled as L, +H, and Ha, on each of the three syllables. Figure 8.4 shows an example of the 'L L+ Ha' surface tonal pattern, and Figure 8.5 shows example APs with two 'L +H Ha' patterns.

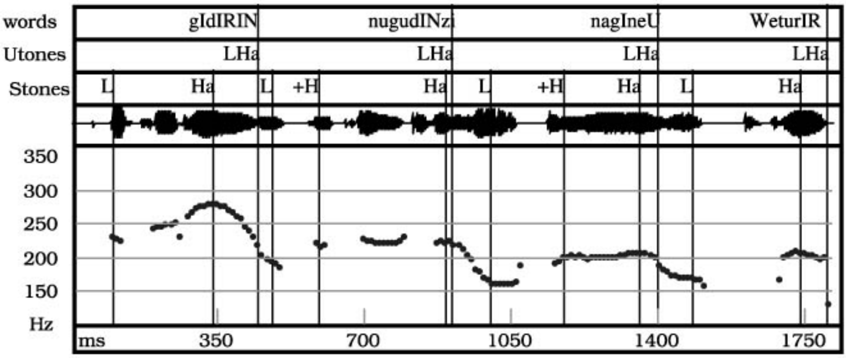


FIGURE 8.5 An example utterance, gIdIRIN 'They-TOP'+'nugudINzi 'whoever'+ nagIneU 'stranger-POSS.'+WeturIR 'clothes-ACC' => 'They, whoever (takes off) the traveller's clothes (first)', illustrating APs with two 'L+H Ha' surface tone patterns (2nd and 3rd APs).

Schematic fo contours of fourteen types of AP realizations and corresponding phonetic tone labels are shown in Figure 8.6. The first row shows AP patterns with a high boundary, Ha, and the second row shows AP patterns with a low boundary, La. The third row shows contours of a long AP where all four underlying tones are realized with either a Ha or La boundary. 'T' in the last contour is either H or L.

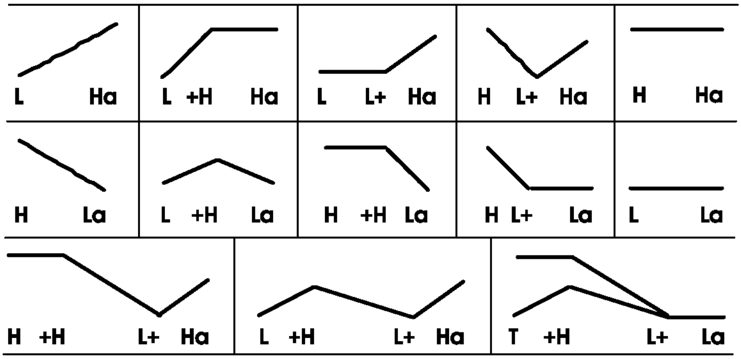


FIGURE 8.6 Schematic fo contours of fourteen tonal patterns of an AP, labelled in tones of the phonetic tone tier.

For the IP boundary tones, the whole tone is placed toward the end of the IP-final syllable aligned with the f0 maximum for H ending boundary tones (i.e., H%, LH%, HLH%, LHLH%) and the f0 minimum for L ending tones (i.e., L%, HL%, LHL%, HLHL%, LHLHL%). For complex boundary tones which include H before the last tone (e.g. HL%, HLH%, LHLH%, LHLHL%), the label '>' should be placed at the f0 peak corresponding to each non-final H tone. Here, '>' can mean an 'early peak' as in English ToBI (i.e., some instances of HL%; see next paragraph), but most of the time it simply indicates the location of H so that it provides information about pitch range. At the moment, it is not clear if complex boundary tones with more than three tones (i.e., LHLH%, HLHL%, LHLHL%) have a distinct meaning of their own other than intensifying the meaning of the less complex tones with two or three tones (e.g. HLHL% intensifies the meaning of HL% or LHL%). More K-ToBI labelled data would be needed to clarify this issue. Until then, all boundary tones should be labelled on the phonetic tone tier.

Currently, the type of IP boundary tone is determined by the f0 shapes on the IP-final syllable. Though this is accurate most of the time, a deviation is sometimes found in news broadcasting where the H tone of HL% is realized on the penultimate syllable of an IP. Park (2003) found a similar phenomenon when an object was postposed after a verb whose boundary tone in the original sentence was HL%, as shown in (1).

(1) gIgE EdisE saDni? => [{gIgE}AP{EdisE saDni}]IP
'that where bought' HL%
'Where did you buy that?'
Postposed object: EdisE saDni, gIgE ? => [{EdisE saDni}{gIgE}]IP
Surface f0 contour: H L%

Park claimed that the H tone on the verb-final syllable ('ni') in the postposed sentence is not an AP final tone, but a part of an IP-final boundary tone, i.e., HL%. The syllable 'ni' is not lengthened (i.e., Break Index 1) and the meaning of the original sentence linked to the HL% boundary is preserved. Importantly, the meaning of the sentence is not preserved if the sentence is produced in two IPs with the first IP ending with a H% tone and the second IP ending with a L% tone. Park called this phenomenon 'tone split' and showed that the H of HL% could be realized earlier than the penult of an IP. She further showed that a tone split can happen to some of the other complex boundary tones beginning with a H tone (e.g. HLHL%), but not those beginning with a L tone (Park 2003). More data need to be examined to see if this phenomenon happens in different constructions or contexts.

The following shows surface realization rules of each boundary tone and its location relative to words and fo contours.

IP-final boundary tones:

L% A level ending or a gently falling boundary tone spread over much of the IP-final syllables. This tone is the most common in stating facts and in declaratives in reading.

H% A rising boundary tone that begins to rise before the IP-final syllable and reaches its peak during the final syllable. Therefore, the rise is earlier than that in LH%. This tone is the most common in seeking information as in yes/no-questions.

LH% A rising boundary tone that is more localized than H%, rising sharply from a valley well within the final syllable. That is, by comparison to H%, this is a sharper, later rise, starting after the onset of the final syllable. This is commonly used for questions, continuation rises, and explanatory endings. It is also used to signal annoyance, irritation, or disbelief (e.g. <gIrEHtaniKa gIrEne!> 'I've already told you so. (Why do you keep asking me?)' or <bEryESE!> '(Did you) throw it out? (I can't believe that!)'). See Figure 8.7 (a) and (b) to compare H% and LH%.

HL% A falling boundary tone that rises *before* the last syllable, and reaches its peak and then falls during the last syllable. Though it seems to be a combination of H% and L%, the H part of this boundary tone is not as high as a simple H% and the L is not as low as a simple L%. This tone is most common in declaratives and wh-questions. It is also commonly used in news broadcasting.

LHL% A rising-falling boundary tone that, unlike HL%, rises *within* the IP-final syllable. The fo peak on H is not as high as that of LH%. It sometimes intensifies the meaning of HL%, but like LH%, it also delivers the meanings of being persuasive, insisting, and confirmative. It is also used to show annoyance or irritation (e.g. <hazima>! 'Don't do it (I told you before)'). See Figure 8.8 (a) and (b) to compare HL% and LHL%.

HLH% A falling-rising boundary tone—a combination of HL% and H% in that the timing of the rise is the same as HL% but followed by a shallow dip and then another rise.

The location of the first H should be marked by '>' above the fo peak. This tone is used when a speaker is confident and expecting listeners' agreement. An example of HLH% is shown in Figures 8.9 and 8.13.

LHLH% A rising-falling-rising boundary tone. The timing of the rise is like LH%. The location of the first H should be marked by '>' above the fo peak. This tone is rare and has a meaning of intensifying some of the LH% meanings, i.e., annoyance, irritation, or disbelief.

HLHL% A falling-rising-falling boundary tone. The timing of the rise is like HL%. The location of the two Hs should be marked by '>' above the fo peaks. This tone is more common than LHLH% but not as common as less complex boundary tones. It sometimes intensifies the meaning of HL%, confirming and insisting on one's opinion, and sometimes, like LHL%, it delivers nagging or persuading meanings.

LHLHL% A rising-falling-rising-falling boundary tone. The timing of the rise is like LH% followed by LHL%. The location of the two Hs should be marked by '>' above the fo peaks. This tone is rare, and its meaning is similar to that of LHL%, but has a more intense meaning of being annoyed.

Figures 8.7–8.9 show examples of IP boundary tones: H% and LH% in Figure 8.7, HL% and LHL% in Figure 8.8, and HLH% in Figure 8.9. In Figures 8.7 and 8.8, a vertical dashed line marks the beginning of the last syllable, '-yo' [jo], showing the timing of the rise with reference to the final syllable.

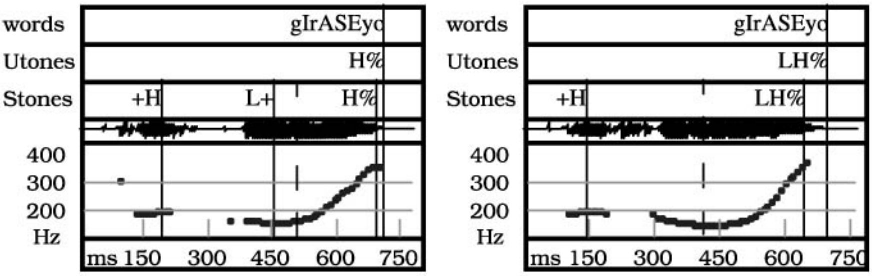


FIGURE 8.7 One-word utterance, gIrASEyo 'Is that so?', with (a) H% and (b) LH%.

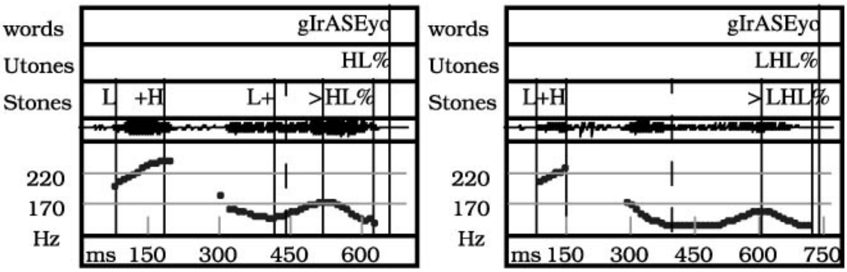


FIGURE 8.8 One-word utterance, gIrASEyo 'Is that so?', with (a) HL% and (b) LHL%.

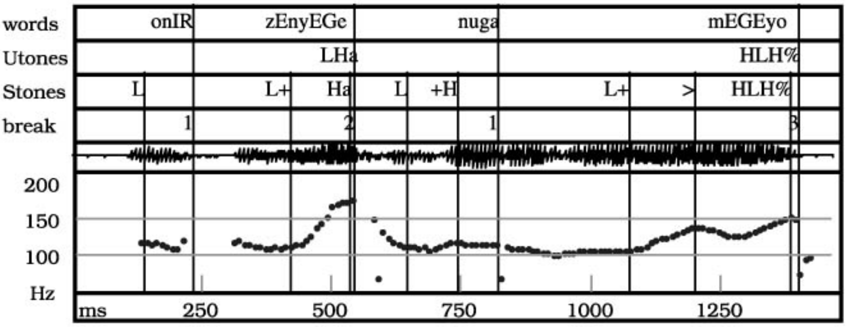


FIGURE 8.9 An example utterance, onIR 'today'+zEnyEGe 'night'+nuga 'who'+ mEGEyo 'eat?' => 'Who is eating tonight?', illustrating HLH%.

Schematic fo contours of eight types of IP boundary tone realizations are shown in Figure 8.10. The first row shows IP boundaries ending with L% and the second row shows those ending with H%. The vertical line shown in each contour marks the beginning of the IP-final syllable. The fo scale is not normalized.

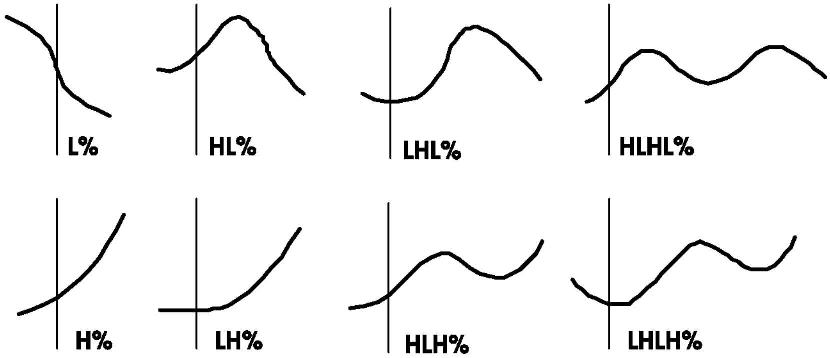


FIGURE 8.10 Schematic fo contours of eight boundary tones of IP.

Finally, for cases of uncertain or underspecified tonal events, for both AP and IP, use the following labels on the phonetic tone tier. Underspecified tone labels should be used when a labeller knows there is a tone, but has not assigned a label yet.

X Underspecified tonal event of a non-AP-final tone. (Tone is there, but the tonal value has yet to be assigned.)

a Underspecified AP-final tone.

% Underspecified IP-final tone.

X? Uncertain of the type of tone, which could be either an AP-final or IP-final boundary tone. (The labeller is not sure which of the two tone types to assign.)

Xa? Uncertain of the type of AP-final boundary tone.

X%? Uncertain of the type of IP-final boundary tone.

(iv) *The break indices tier*: break indices represent the degree of juncture perceived between each pair of words and between the final word and the silence at the end of the utterance. They are to be marked after all words that have been transcribed in the word tier. All junctures—including those after fragments and filled pauses—must be assigned an explicit break index value. Values for the break index are chosen from the following set. An example of a Break Index (BI) 0 is shown in Figure 8.11, and those of BIs 1, 2, 3 are shown in Figure 8.9. (The break index tier is named 'break'.)

0 For cases of clear phonetic marks of 'clitic' groups; e.g. application of vowel coalescence rules. Also for cases of 'incomplete nouns' (monosyllabic nouns which, though separated by spaces, are seldom used by themselves but need a modifier; e.g. <su> 'way', <de> 'place', <gED> 'thing').

1 For phrase-internal 'word' boundaries which are not marked by such cliticization phenomena and can be pronounced independently.

2 For cases of a minimal phrasal disjuncture, with no strong subjective sense of pause—that is, a sense of phrase edge of the type that is typically associated with the Accentual Phrase final tone.

3 For cases of a strong phrasal disjuncture, with a strong subjective sense of pause (whether it be an objectively visible pause or only the 'virtual pause' cued by final lengthening)—that is, a sense of phrase break of the type that is typically associated with the boundary tone of an Intonation Phrase.

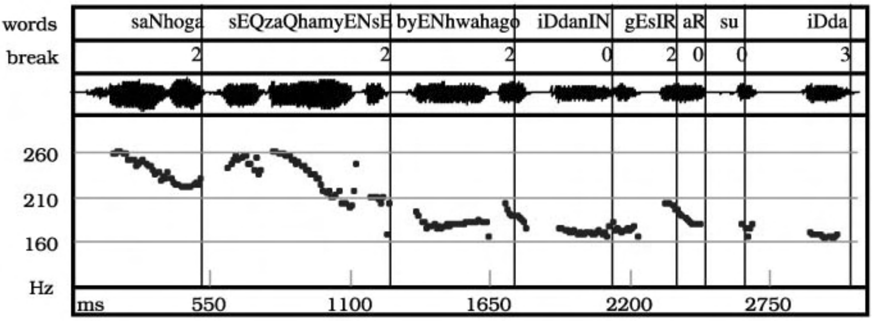


FIGURE 8.11 An example utterance, saNhoga ‘coral-NOM.’+sEQzaQhamyENsE ‘growing’+byENhwahago iDdanIN ‘change-prog.-rel.’+gEsIR ‘thing-ACC.’+aR su iDda ‘to see’ => ‘(We) can see that the coral is changing while growing’, illustrating BI 0.

Note that while the AP and IP are defined in the prosodic model by tonal markings, the break index value indicates the labeller’s subjective sense of disjuncture and not simply the juncture that typifies the apparent tones. Thus, the break index tier markings for break index levels 2 and 3 are not made completely redundant by the tone tier markings. In cases of mismatch, the break index number should be chosen following the perceived juncture rather than the tones, and it should be flagged with the diacritic ‘m’. Though it is logically possible to have 3m, ‘3m’ with an AP tone will not be found in data because an AP-final tone, L or H, is one of the IP-final tones. That is, there would be no such case as ‘3-like juncture with an AP-final tone, not an IP-final tone’. Definitions of 1m and 2m are given below. Figure 8.12 shows an example utterance of 1m (1-like juncture with Ha), and Figure 8.13 shows an example utterance of 2m (2-like juncture with no AP-final tone).

1m A disjuncture that typically would correspond to a phrase-medial word boundary, but is marked by the tonal pattern of an AP.

2m A medium strength disjuncture that typically would be marked by the tonal pattern of the AP, but has no tonal markings, or has the tonal markings of an IP edge.

Transcriber uncertainty about break-index strength is indicated with a minus (‘–’) diacritic affixed directly to the right of the higher break index (e.g. ‘1–’ to indicate uncertainty between ‘0’ and ‘1’; ‘2–’ to indicate uncertainty between ‘1’ and ‘2’). An example of BI 2– is shown in Figures 8.12 and 8.13. Note that since the ‘m’ diacritic suggests certainty about the break index

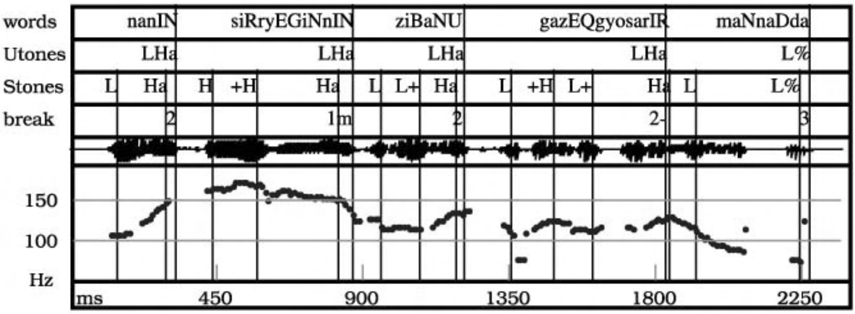


Figure 8.12 An example utterance, nanIN 'I-top'+siRryEGiNnIN 'powerful'+ziBaNU 'family's'+gazEQgyosarIR 'tutor-Acc.'+maNnaDda 'met' => 'I met the tutor of a powerful family', illustrating BIs 1m, 2–, 2, and 3.

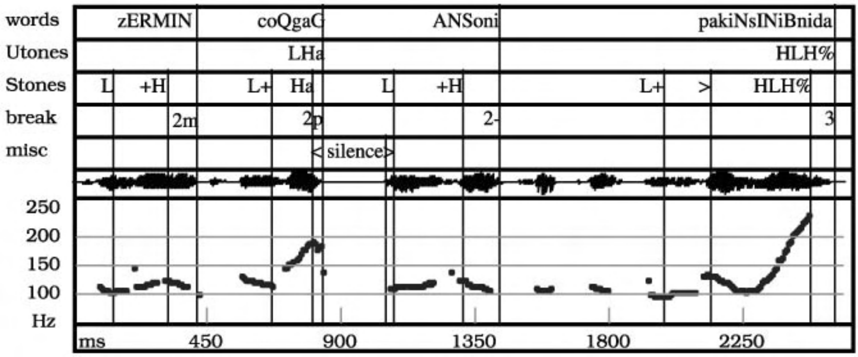


Figure 8.13 An example utterance, zERMIN 'young'+coQgaG 'bachelor'+ANSoni 'Anthony'+pakiNsINiBnida 'Parkinson-be' => '(The firmly standing guard is) the young bachelor, Anthony Parkinson', illustrating BIs 2m, 2–, 2p, and 3.

analysis in the face of conflicting tonal evidence, the '–' diacritic should not be used together with 'm'. Finally, a hesitational pause or disfluency after a word is labelled with a 'p' diacritic affixed directly to the right of the break index—e.g. '1p' for abrupt cut offs after or in the middle of a word and '2p' for prolongation of an AP-final syllable, but not meant to be IP final. An example of '2p' is shown in Figure 8.13.

(v) *The miscellaneous tier*: the miscellaneous tier will be used for any comments on speech (e.g. silence, audible breaths, laughter, disfluencies) desired by particular transcription groups. The only conventions K-ToBI specifies for this tier are that events that cover some clearly specifiable interval (such as breaths, silence, or laughter) be labelled by the < . . . . > pair, aligned with both their temporal beginnings and ends. Event labels are written only before '>', as illustrated below. An example is shown in Figure 8.13. (The miscellaneous tier is named 'misc'.)

<	beginning of an interval (e.g. laughter)
laughter>	end of a period of laughter

8.4. LABELLER AGREEMENT

Analysis of labeller agreement and consistency in the transcription of Korean prosody using K-ToBI conventions described above has been performed based on twenty utterances representing five different types of speech (i.e., news, interview, text reading, story reading, and soap-opera) produced by eighteen speakers (Jun *et al.* 2000). Tones and break indices were transcribed by twenty-one labellers differing in their levels of experience with K-ToBI (five experts, six familiar with K-ToBI, five familiar with the British intonation model but new to K-ToBI, and five beginners). Following the stringent metric used for English ToBI evaluation (Silverman *et al.* 1992; Pitrelli *et al.* 1994), consistency was measured in terms of the number of transcriber pairs agreeing on the labelling of each particular word.

The results show that for tonal transcriptions of the 32,130 transcriber-pair-words, agreement was 77.3 per cent for the type of boundaries at the end of each word (i.e., word, AP, or IP), 77.5 per cent for AP boundaries, and 90.9 per cent for IP boundaries. For experts, agreement was 81.6 per cent, 81.9 per cent, and 90.9 per cent, respectively. For agreement on the surface realization of AP tones, agreement for the word-initial tone (L, H, no tone) was 82.3 per cent for all labellers and 90.7 per cent for experts, and that for the word-final tone (La, Ha, L, H, +H, L+, no tone) was 74.9 per cent for all labellers and 82.0 per cent for experts. About 11 per cent of AP-initial tones were not predictable from the segmental information (i.e., H when aspirated or tense consonants, and L otherwise), and about 16 per cent of AP-final tones were not rising but falling, i.e., La.

Agreement for the whole tonal pattern for each word, however, was only about 36 per cent for all labellers and 52 per cent for experts. This low agreement seems to be due to the nature of the tonal pattern. That is, there are fourteen possible AP tonal patterns whose surface variations do not seem to be meaningful or phonological. Furthermore, there is a gross similarity among some of the tonal patterns. For example, rising tonal patterns such as LH, LLH, and LHH are very similar to one another, and falling patterns such as LHLL and LHL are also very similar to each other.

For break indices excluding sentence final BIs, the agreement score among all labellers was 59 per cent for exact matching, 69 per cent when relaxing for the presence/absence of diacritics, and 99 per cent when relaxing within +/ – 1 level. The agreement score among experts was 65.5 per cent, 77.1 per cent, and 99.0 per cent, respectively. The results are, in general, close to those for English ToBI (66.6 per cent, 70.4 per cent, 92.5 per cent, respectively; Pitrelli *et al.* 1994) with somewhat lower results for exact matching (59 per cent vs. 66.6 per cent). Relaxing to the +/– 1 level criterion results in higher agreement for Korean than for English most likely because Korean has four levels (0–3) of BIs, while English has five (0–4).

The results confirmed that the conventions of K-ToBI are adequate, easy to learn, and can be reliably used for research in Korean prosody and for large-scale prosodic annotation in speech databases. More data need to be transcribed to confirm these statistics and to find out the frequency and the function of the surface realization of AP tones and IP boundary tones as well as the detailed alignment of tones.

8.5. CONCLUSION

In this chapter, we have presented an overview of Korean intonational structure and the most updated version of K-ToBI labelling conventions. Korean has two prosodic units defined by intonation, Accentual Phrase and Intonation Phrase, and there are at least fourteen surface tonal patterns of an AP and nine types of IP boundary tones. The main revision of the current K-ToBI is splitting the tone tier into two tiers: the phonetic tone tier, labelling the surface tonal patterns of AP and IP boundary tones, and the phonological tone tier, labelling the tones marking the boundary of prosodic units, AP and IP. Labeller agreement data showed lower agreement percentage for surface tonal patterns than for tones marking prosodic boundaries, supporting the division between phonetic and phonological tones.

Korean has four degrees of juncture, one within a word and three between words. A diacritic 'm' is used after a break index when the degree of juncture corresponding to the break index does not match the expected tonal pattern. Agreement on the break index labelling was found to be similar to that of English ToBI. The agreement data showed that K-ToBI conventions are adequate and reliable. Transcribing more data using K-ToBI is needed to find distinctive meanings of tones and their realizations and to further improve K-ToBI conventions.

The current K-ToBI version (Version 3) of the K-ToBI manual is available on the website (http://www.linguistics.ucla.edu/people/jun/sunah.htm) and also on the UCLA Phonetics Lab website (http://www.linguistics.ucla.edu/faciliti/uclaplab.html).

APPENDIX A: SUMMARY OF K-ToBI LABELS (TONES AND BREAK INDICES)

Tier	Label	Description
Phonological tones	LHa	AP final boundary
	L%	IP final low boundary tone. Declarative
	H%	IP final high boundary tone. Interrogative
	LH%	IP final rising boundary tone. Interrog., Cont. rising
	HL%	IP final falling boundary tone. Declarative, Wh-Q
	LHL%	IP final rising-falling boundary tone. Insisting
	HLH%	IP final falling-rising boundary tone. Declarative.
	LHLH%	IP final rising-falling-rising boundary tone. Insisting, Declarative, Intensifying some of LH% meaning
	HLHL%	IP final falling-rising-falling boundary tone. Intensifying some of HL% meaning
	LHLHL%	IP final rising-falling-rising-falling boundary tone.
Phonetic tones	L	AP initial low tone (default for APs beginning with sonorants and lenis obstruents).
	H	AP initial high tone (default for APs beginning with aspirated or tense obstruents).
	+H	High tone on the second syllable of an AP. Optional
	L+	Low tone on the penult of an AP. Optional
	La	AP final low boundary tone.
	Ha	AP final high boundary tone.
	L%, H%, LH%, HL%, LHL%, HLH%, LHLH%, HLHL%, LHLHL%	— same as the IP tones on the phonological tone tier.
Break index	0	Reduced/cliticized word boundary.
	1	AP medial word boundary
	2	AP boundary
	3	IP boundary

m	Diacritic for mismatch. 1m: AP-med juncture with AP final tone; 2m: AP-final juncture with no AP or IP tone
-	Uncertainty between two adjacent break indices, ex. 3-: uncertain between BI2 and BI3.
p	Diacritic for disfluency or hesitation pause
>	Delayed peak.

APPENDIX B

Romanization Convention[a]

1. *Consonants*

Hangul	[IPA]	Roman letters	
		Onset	Coda
ㅂ	[p]	b	B
ㄷ	[t]	d	D
ㄱ	[k]	g	G
ㅈ	[tʃ]	z	D/z
ㅍ	[p^{h}]	p	B/p
ㅌ	[t^{h}]	t	D/t
ㅋ	[k^{h}]	k	G/k
ㅊ	[tʃ^{h}]	c	D/c
ㅃ	[p*]	P	—
ㄸ	[t*]	T	—
ㄲ	[k*]	K	G/K
ㅉ	[tʃ*]	C	—
ㅅ	[s]	s	D/s
ㅆ	[s*]	S	D/S
ㅎ	[h]	h	H
ㄹ	[l/ɾ]	r	R
ㅁ	[m]	m	M
ㄴ	[n]	n	N
ㅇ	[ŋ]	—	Q

2. *Vowels*

Hangul	[IPA]	Roman letters
아	[ɑ]	a
어	[ɤ]	E
오	[o]	o
우	[u]	u
으	[ɨ]	I
이	[i]	i
에	[e/ɛ]	e
애	[ɛ]	A
의	[ɯi]	U
야	[ja]	ya
여	[jə]	yE
요	[jo]	yo
유	[ju]	yu
예	[je/jɛ]	ye
얘	[jɛ]	yA
와	[wa]	Wa
워	[wə]	wE
웨	[we/wɛ]	we
외	[we/wɛ]	We
왜	[we/wɛ]	WA
위	[wi/ɥi]	wi

[a] Some of the Coda consonants are represented with two Roman letters separated by a slash. In this case, the first letter is for the Coda consonant followed by an Onset consonant, and the second letter is for the Coda consonant followed by a vowel, i.e., no Onset. This division reflects a neutralized Coda before a syllable with an Onset consonant (e.g. KoDziB 'a flower store') and a resyllabified Coda into an Onset before a Vowel initial syllable (e.g. Koci 'a flower-NOM.'). Three tense consonants, [p*], [t*], and [tʃ*] occur as onsets only, and [ŋ] as a coda only.

REFERENCES

Arvaniti, A., and Baltazani, M. (this volume Ch. 4), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora'.

Beckman, M. E., and Ayers, G. (1994), 'Guidelines for ToBI Labeling'. On line ms Ohio State University. <http://www.ling.ohio-state.edu/~tobi/ame_tobi>.

——, and Hirschberg, J. (1994), 'The ToBI Annotation Conventions', ms Ohio State University.

——, and Jun, S.-A. (1996), 'K-ToBI (Korean ToBI) Labeling Convention' Version 2, ms Ohio State University and UCLA. Manuscript is available at <http://www.linguistics.ucla.edu/people/jun/sunah.htm>.

——, and Pierrehumbert, J. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–309.

Cho, S.-B. (1967), *A Phonological Study of Korean with a Historical Analysis* (Uppsala: Universitet).

Cho, T., and Keating, P. (2001), 'Articulatory Strengthening at the Onset of Prosodic Domains in Korean', *Journal of Phonetics*, 28: 155–90.

de Jong, K. (1989), 'Initial Tones and Prominence in Seoul Korean', paper presented at the 117th meeting of the Acoustical Society of America, Syracuse, NY, 22–6 May; and published in the *Ohio State University Working Papers in Linguistics*, No. 43, 1–14 (1994).

—— (2000), 'Attention Modulation and the Formal Properties of Stress Systems', in A. Okrent and J. P. Boyle (eds.), *Chicago Linguistic Society*, 36: 71–92.

Fougeron, C., and Jun, S.-A. (1998), 'Rate Effects on French Intonation: Phonetic Realization and Prosodic Organization', *Journal of Phonetics*, 26/1: 45–70.

Halle, M., and Stevens, K. (1971), 'A Note on Laryngeal Features', *Quarterly Progress Report*, 101: 198–212 (Cambridge, MA: Research Laboratory of Electronics, MIT).

Hayes, B. (1989), 'The Prosodic Hierarchy in Meter', in P. Kiparsky, and G. Youmans (eds.), *Perspectives on Meter* (New York: Academic Press), 203–60.

Hirst, D., and Di Cristo, A. (1984), 'French Intonation: A Parametric Approach', *Die Neueren Sprachen*, 83/5: 554–69.

——, —— (1998) (eds.), *Intonation Systems: Survey of Twenty Languages* (Cambridge: Cambridge University Press).

Jun, S.-A. (1989), 'The Accentual Pattern and Prosody of Chonnam Dialect of Korean', in S. Kuno, I.-H. Lee, J. Whitman, S.-Y. Bak, Y.-S. Yang, and Y.-J. Kim (eds.), *Harvard Studies in Korean Linguistics* III (Cambridge, MA: Harvard University), 89–100.

—— (1990), 'The Prosodic Structure of Korean—in Terms of Voicing', in E.-J. Baek (ed.), *Proceedings of the Seventh International Conference on Korean Linguistics* (University of Toronto Press), 7: 87–104.

—— (1993), 'The Phonetics and Phonology of Korean Prosody', Ph.D. dissertation (Ohio State University). [Published in 1996 by New York: Garland].

—— (1995*a*), 'A Phonetic Study of Stress in Korean', poster presented at the 130th meeting of the Acoustical Society of America, St Louis, MO, *JASA* 98 (5–2): 2898.

—— (1995*b*), 'Asymmetrical Prosodic Effects on the Laryngeal Gesture in Korean' in B. Connell and A. Arvaniti (eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology* (Cambridge, UK: Cambridge University Press), 4: 235–53.

—— (1996), 'Influence of Microprosody on Macroprosody: A Case of Phrase Initial Strengthening', *UCLA Working Papers in Phonetics*, 92: 97–116.

—— (1998), 'The Accentual Phrase in the Korean Prosodic Hierarchy', *Phonology*, 15/2: 189–226.

—— (2000), 'K-ToBI (Korean ToBI) Labeling Conventions: Version 3', *The Korean Journal of Speech Sciences*, 7/1: 143–69 [Version 3.1 is published in *UCLA Working Papers in Phonetics*, 99: 149–73].

——, and Fougeron, C. (1995), 'The Accentual Phrase and the Prosodic Structure of French', in *Proceedings of XIIIth International Congress of Phonetic Sciences* (Stockholm, Sweden), 2: 722–5.

——, —— (1997), 'A Phonological Model of French Intonation', poster presented at the *ESCA Workshop on Intonation*, (Athens, Greece) 18–20 September.

——, —— (2000), 'A Phonological Model of French Intonation', in A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology* (Dordrecht, Netherlands: Kluwer Academic Publishers), 209–42.

——, —— (2002), 'Realizations of the Accentual Phrase in French Intonation', in J. Hualde (ed.), *Intonation in the Romance Languages*, special issue of *Probus*, 14: 147–72.

——, Lee, S.-H., Kim, K., and Lee, Y.-J. (2000), 'Labeler Agreement in Transcribing Korean Intonation with K-ToBI', in *Proceedings of the 2000 International Conference on Spoken Language Processing* (Beijing, China), 3: 211–14.

——, and Oh, M. (1996), 'A Prosodic Analysis of Three Types of Wh-phrases in Korean', *Language and Speech*, 39/1: 37–61.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2004), 'Domain-initial Strengthening in Four Languages', in *Laboratory Phonology VI* (Cambridge: Cambridge University Press), 143–61.

Kim, J.-J., Lee, S.-H., Ko, H.-J., Lee, Y.-J., Kim, S.-H., and Lee, J.-Ch. (1997), 'An Analysis of some Prosodic Aspects of Korean Utterances Using K-ToBI Labelling System', in the *Proceedings of ICSP '97* (Seoul, Korea), 87–92.

Koo, H.-S. (1986), 'An Experimental Acoustic Study of the Phonetics of Intonation in Standard Korean', Ph.D. dissertation (University of Texas at Austin) (Seoul: Hanshin Publishing).

Korea Telecom Research and Development Group Report (1996), *A Study of Korean Prosody and Discourse for the Development of Speech Synthesis/Recognition System.* [In Korean].

Ladd, D. R. (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

Lee, H.-B. (1964), 'A Study of Korean (Seoul) Intonation', thesis, University College London.

—— (1974), 'Rhythm and Intonation of Seoul Korean [in Korean]', *Ehag YEngu* (*Language Research*) (Language Research Center, Seoul National University), 10/2: 415–25.

Lee, H.-J. (1999), 'Tonal Realization and Implementation of the Accentual Phrase in Seoul Korean', MA thesis (University of California, Los Angeles).

——, and Kim, H.-S. (1997), 'Phonetic Realization of Seoul Korean Accentual Phrase', *Harvard Studies in Korean Linguistics*, 7: 153–61.

Lee, H.-Y. (1990), 'The Structure of Korean Prosody', dissertation (University of London) (Seoul: Hanshin Publishing).

—— (1997), *GugE Unyulron* (Korean Prosody), Korean Research Center Series 65.

Lee, S.-H. (1989), 'Intonational Domains of the Seoul Dialect of Korean', paper presented at the 117th meeting of the Acoustical Society of America, Syracuse, NY, 22–6 May. Abstract in *Journal of the Acoustical Society of America*, 85, suppl. 1, S99.

Lim, B.-J. (2001), 'The Role of Syllable Weight and Position on Prominence in Korean', *Japanese Korean Linguistics*, 9: 139–50 (Stanford, CSLI).

——, and de Jong, K. (1999), 'Tonal Alignment in Standard Korean: The Case of Younger Generation', paper presented at the Western Conference on Linguistics, the University of Texas, El Paso, TX.

Martin, S. E. (1954), *Korean Morphophonemics.* William Dwight Whitney Linguistic Series (Baltimore, MD: Linguistic Society of America).

Nespor, M., and Vogel, I. (1986), *Prosodic Phonology* (Dordrecht: Foris).

O'Connor, J. D., and Arnold, G. F. (1973), *Intonation of Colloquial English* (2nd edn., London: Longman).

Park, M.-J. (2003), 'The Meaning of Korean Prosodic Boundary Tones', Ph.D. dissertation (University of California, Los Angeles).

Pierrehumbert, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

——, and Beckman, M. E. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

Pitrelli, J., Beckman, M. E., and Hirschberg, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework', in *Proceedings of the 1994 International Conference on Spoken Language Processing* (Yokohama, Japan), 1: 123–6.

Polivanov, V. E. (1936), 'Zur Frage der Betonungsfunktionen', in *Etudes Dediees au Quatrieme Congres de Linguistes, TCLP*, 6: 75–81.

Schafer, A., and Jun, S.-A. (2002), 'Effects of Accentual Phrasing on Adjective Interpretation in Korean', in M. Nakayama (ed.), *East Asian Language Processing*, Stanford, CSLI, 223–55.

Selkirk, E. O. (1984), *Phonology and Syntax: The Relation between Sound and Structure* (Cambridge, MA: MIT Press).

Seong, C.-J. (1995), 'The Experimental Phonetic Study of the Standard Current Korean Speech Rhythm—with respect to its temporal structure', Ph.D. dissertation (Seoul National University) (Seoul: Hanshin Publishing).

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992), 'ToBI: A Standard for Labeling English Prosody', in *Proceedings of the 1992 International Conference on Spoken Language Processing* (Banff, Canada), 2: 867–70.

Trubetzkoy, N. S. (1939), *Grundzuge der Phonologie. TCLP* 7.

Venditti, J. (1995), 'Japanese ToBI Labeling Guidelines', ms Ohio State University.

—— (this volume Ch. 7), 'The J-ToBI model of Japanese Intonation'.

9

Towards a Pan-Mandarin System for Prosodic Transcription

Shu-hui Peng, Marjorie K. M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, and Mary E. Beckman

9.1. INTRODUCTION

This chapter describes the initial stages of development of a Pan-Mandarin ToBI system. We will review the salient prosodic characteristics of Mandarin, with particular attention to the range of variability within a common structural core, and then propose an initial codification of conventions for marking prosodic structure in two standard varieties and one regional variety of the language. Before we begin, however, we must at least pose two important prior questions. First, is a Pan-Mandarin ToBI feasible? Second, if feasible, is it desirable?

The first question arises because of the sheer size of the language. Mandarin is spoken as a first or second language by well over a billion people. Furthermore, it has had a long history of being spoken over large geographical areas as a standard language or as a local language. As a result, Mandarin encompasses many different varieties, including three national standards (Guoyu in Taiwan, Putonghua in Mainland China, and Huayu in Singapore) as well as many regional varieties, such as Rugaohua, discussed in Section 9.2.5. Furthermore, a large population of Mandarin speakers speaks more than one variety of Mandarin. Many others speak a variety of Mandarin and some non-Mandarin Chinese or other language variety. Code-switching often occurs, at many different levels, including prosody. The enormous geographical range and the interaction with different substrate languages at the edges of its range lead to variability in syntax, in lexicon, and especially in phonology, including the phonology of tone, stress, and prosodic grouping. The short answer to the first question, therefore, is that we cannot know until we try.

At the same time, we think that this variability makes Mandarin an ideal test bed for developing the ToBI framework into a more powerful descriptive tool.

While a Pan-Mandarin ToBI system may be too ambitious, this is our goal. The current Mandarin ToBI system is being designed to describe the prosodic structure and inventory of tones and other structure-marking elements in Putonghua, Guoyu, and several regional varieties. Eventually, we hope that it can be extended to cover all varieties. In developing the system, then, we need to specify how to expand it incrementally to accommodate our expanding knowledge of regional varieties as well as of the standard varieties. With the right accommodation, the system should also be able to describe interactions (such as code-switching events) between different varieties of Mandarin and perhaps between Mandarin and other varieties of Chinese (and other languages) in different social contexts. The accommodation to variation and code-switching is of course a more general issue for the ToBI framework (see e.g. Bruce this volume Ch. 15; Grice *et al.* this volume Ch. 13). Thus, a general answer to the second prior question is a resounding affirmative.

We can also answer the second question more specifically in response to the aims of this book. The Mandarin varieties of Chinese are a particularly fruitful source of data for typological comparison across prosodic systems. There are several reasons for this. First, all varieties of Mandarin have lexical tone contrasts, and all show a fairly dense specification of tone (albeit none as dense as in Cantonese—see Wong *et al.*, this volume Ch. 10). At the same time, many Mandarin varieties have metrical structure reminiscent of the stress systems seen in the Germanic languages, with tone specification constrained by syllable prominence at both the lexical and the phrasal levels. Thus, just as Cantonese nicely illustrates the interaction between lexical tone and the 'purely intonational' use of tones as pragmatic morphemes that mark phrase edges, Mandarin illustrates the interaction between lexical tone and the 'accentual' use of tone as a stress marker.

Mandarin is a fruitful source of data for typological comparison also because there is enough variability just within Mandarin itself to look at the effects of minimal differences in just one aspect of the prosodic system while keeping the lexicon and the rest of the prosodic core intact. For example, there is variability across varieties in the incidence and distribution of unstressed (toneless) syllables that is fully comparable to that seen in the Germanic languages. Just as Swedish differs markedly from English and German in the distribution of 'de-accented' words in running speech, Guoyu (the standard Mandarin of Taiwan) differs markedly from Putonghua (the standard of the People's Republic of China) in the distribution of 'neutral tone' syllables (Tai 1978).

Another point of difference across Mandarin varieties lies in the evidence for categorically marked prosodic units above the syllable. For example, all varieties of Mandarin have 'tone sandhi'—i.e. phonologically- or lexically-conditioned tonal alternations that are vaguely reminiscent of the phenomena called 'tone sandhi' in the Wu varieties of Chinese, such as Shanghainese (Jin 1985), Danyang (Chan 1991) and Wuxi (Chan and Ren 1989). However, unlike in the Wu varieties, the relationship between tone sandhi and prosodic structure in standard varieties of Mandarin is quite controversial. Some earlier accounts emphasized the blocking of tone sandhi at certain syntactic constituent edges (e.g. Cheng, C.-C. 1973), which might suggest a Wu-like phonological phrase smaller than the intonational phrase as the domain of tone sandhi (see e.g. Liao 1994). This characterization is probably accurate for tone sandhi in the regional varieties that border on the Wu-speaking regions of China. However, for the standard varieties, other more recent accounts emphasize rhythmic constraints, and prominence alternations within the 'foot' (e.g. Shih 1986, 1997; Zhang 1988; Hsiao 1991). In short, there are no unifying commonalities such as those that define the 'tone sandhi group' as a shared prosodic unit in all Wu varieties despite the large and salient differences in details of tone shape and initial- versus final-syllable dominance. We might say that the phonological characterization of delimitative markers for any shared core of higher-level prosodic groupings across varieties is complicated by the relationship in the standard varieties between tone, on the one hand, and stress or 'foot-level' rhythmic grouping, on the other.

In this chapter, then, we will briefly describe the salient characteristics of the Mandarin varieties of Chinese that a ToBI annotation system should capture. We will give due emphasis to comparison across different varieties of Mandarin, in addition to the obvious points of reference to other non-Mandarin varieties of Chinese and to other languages that have been described within the ToBI framework. We will also outline how the ToBI framework has been adapted at two different sites to develop annotation conventions for tagging spoken language databases. Since the research focus differs somewhat between the two sites, with one site working to build the requisite infrastructure for speech synthesis and recognition, and the other site developing corpora to explore sociolinguistic variation in prosody across different varieties and styles, Mandarin ToBI also provides a useful illustration of how basic research in prosodic typology can inform other areas of linguistics and speech science.

9.2. SALIENT CHARACTERISTICS OF MANDARIN

9.2.1. *The scope of Mandarin*

As noted above, one of the characteristics of Mandarin that is especially challenging for the ToBI framework is its size. Some 70 percent of the

Chinese-speaking population of the world are speakers of Mandarin. There are more than 800 million native speakers of Mandarin just in the People's Republic of China, and at least half that many speakers of Mandarin as a second language (see Grimes 1996). Mandarin is spoken in all regions of China north of the Yangtze River and in large regions south of it in south-western China. This geographical distribution is three-quarters of the entire primarily Chinese-speaking territory of the PRC (Yuan 1960, 1989). Thus, the Mandarin-speaking region of mainland China is large enough to encompass considerable variability. Yuan (1960, 1989) classifies Mandarin into four main varieties based in part on geography and in part on phonological differences: Northern Mandarin, North-western Mandarin, South-western Mandarin, and Jianghuai Mandarin. Li, R. (1985, 1989*a*, *b*) designates the Chinese spoken in Shanxi Province in northern China as non-Mandarin, and then subdivides the rest of Yuan's Mandarin into seven major regional varieties based on modern reflexes of the so-called 'entering tone' (*rusheng*). Neither of these classifications includes the variability that has emerged among speakers of Mandarin in other parts of the world, such as Taiwan, Singapore, Indonesia, Thailand and other parts of Southeast Asia, the United Kingdom, North America, and South Africa.

Across the Mandarin-speaking world, Mandarin is in contact with many other varieties of Chinese, as well as of non-Chinese languages. This contact increases the variability even more. For example, the Jianghuai varieties share many phonological features with the neighbouring Wu varieties of Chinese. They have even been described as being underlyingly Wu with a Mandarin superstratum (Ting 1966).

Even discounting such variability across regional varieties, Mandarin is still a very heterogeneous language. Whether speaking Putonghua as a first or a second language, speakers of this standard variety will show regional features and differences associated with age and social class. Even the prescriptive Putonghua of trained broadcasters differs from the 'original' source of the standard, Beijinghua (Hu 1987). And Guoyu, the standard Mandarin of Taiwan, shows features suggesting a Min substratum. This is not surprising, since Taiwanese, the native language of somewhere between 70 percent and 80 percent of Taiwan Chinese, is a Min variety (see Cheng, R.-L. 1985; Feifel 1994). Cheng, R.-L. (1985) also observes that, among the especially influential and economically powerful non-Taiwanese speakers of Mandarin in Taiwan there are many Wu speakers, including the former ruling family. Thus, as a result of contact with Taiwanese, Wu, non-Beijing varieties of Mandarin, and so forth, these two standards today—Putonghua and Guoyu—are quite different. Tai (1976, 1977, 1978), for example, systematically documented phonological, syntactic, lexical, and stylistic differences between Putonghua and Guoyu that already reflected significant changes after only the first quarter-century of political separation. Thus, even if we limited Mandarin ToBI to dealing only with 'standard' Mandarin, there would be much variability to cover.

9.2.2. *Lexical tone*

Lexical tone is another salient characteristic of modern Mandarin. As in other varieties of Chinese, every syllable in a Mandarin utterance can be identified as a free or a bound morpheme. And almost every syllable—even one that is a bound morpheme—is lexically specified for one of several phonemically contrasting tones. Tone specification is so prevalent in the lexicon that the exceptional syllables are traditionally described as having a tonal specification—for 'neutral' tone (see Section 9.2.3).

The number of tones in contrast is often four, as in the standard varieties, although it can be as few as three, as in the Yantai and Qingdao varieties spoken in Shandong Province (see Qian 1982, for Yantai; and Qingdaoshi Shizhi Bangongshi 1997, for Qingdao). Although there are always exceptions, due to 'dialect borrowing' at various stages in the history of the language, it is often possible to identify a fairly simple set of correspondences between cognate forms. However, even when two varieties show such simple correspondences, they can differ markedly in the phonetic tone shapes for cognate morphemes in citation form. Moreover, even between the standard varieties, we see divergence in phonetic shape. Because of this variability, contemporary linguists often refer to the lexical tones as 'Tone 1', 'Tone 2', 'Tone 3' and 'Tone 4' rather than with more descriptive terms such as 'high tone', 'rising tone', and so on, and we will often adopt this terminology in talking about the standard varieties, where the cognate sets are quite well behaved. (Old Mandarin also had four tones, and classically educated linguists sometimes use traditional, philological terms, such as *yin-ping* versus *yang-ping* for Putonghua Tone 1 and Tone 2. However, the correspondence between the four tones of Old Mandarin and those of the modern standard varieties is not simple—cf. Stimson (1966), for example, for a discussion of the modern reflexes in Beijing Mandarin. Hence, we will resort to these philological terms only for describing the tones of 'checked' *rusheng* syllables in regional varieties that retain vestiges of the syllable-final stops of Middle Chinese.)

The four tones in standard Mandarin are described variously by different linguists. Some of these differences are due merely to the phonological framework assumed. Others reflect real contextually conditioned variation. Chao (1948, 1968) lists the citation tone shapes as 'high-level' for Tone 1, 'high-rising' for Tone 2, 'low-dipping' for Tone 3, and 'high-falling' for Tone 4. He further notes that in non-citation context, the third tone surfaces as 'high-rising' before another third tone in tone sandhi contexts, and as 'low-falling' elsewhere. Kratochvil (1968), by contrast, lists the four tones as 'high', 'rising', 'low', and 'falling'—a phonemicization that is amenable to a two-level Autosegmental analysis as H, LH, L, and HL. Only in the phonetic description does he identify the third tone as having a dipping pitch contour in citation forms.

Table 9.1 shows a traditional alphabetic phonetic representation in terms of Chao's (1930/1980) tone numbers (with five levels), taking Chao's 'elsewhere' shape as the basic form for Tone 3, as in Shih (1986, 1988). Note that Shih's decision regarding the 'basic' shape for Tone 3 was intended specifically for Guoyu. This is in keeping with Tai's (1978) observation that even the citation form of Tone 3 tends to be produced as [21] in Guoyu, rather than with the full dipping tone that Chao chose as the 'basic' citation form shape. Both Tai and Shih, then, assume that the Guoyu tone system differs from the Putonghua system primarily in the distribution of the Tone 3 allotones. (However, see Hartman (1944) who describes the dipping variant of Tone 3 only for 'loud-stressed syllables occurring finally', and he was writing before the emergence of Guoyu as an independent standard.)

More recent work has uncovered other differences, which may have emerged only in the quarter century since Tai's (1978) treatise. A corpus of productions by a representative younger speaker described in Fon and Chiang (1999) suggests two differences. First, whereas in Putonghua, a Tone 3 syllable in its 'full' [214] form is the longest of all citation form types, the corresponding 'dipping' variant of Tone 3 in Guoyu is shorter than Tone 2. Thus, the relationship between the falling and the dipping variant in Guoyu cannot be ascribed to tone truncation. Also, Tone 2, which in Putonghua is typically a rising tone in citation form, in Taiwan Mandarin is another

TABLE 9.1 Contrastive tones in standard Putonghua Mandarin

Tone	Description	Transcriptions		Example	Pinyin
Tone 1	high level	55	H	ba55 'a scar'	bā
Tone 2	high rising	35	LH	ba35 'to uproot'	bá
Tone 3	low falling	21	L	ba21 'a target'	bǎ
Tone 4	high falling	51	HL	ba51 'a dam'	bà
Neutral tone				ba '(particle)'	ba

Table 9.2 Contrastive tones in Rugaohua Mandarin

Tone	Description	Transcriptions		Example	
Tone 1	falling	41	HL	ba^{41}	'a scar'
Tone 2	rising	35	MH	pa^{35}	'to crawl'
Tone 3	low (falling-rising)	323	MLM	ba^{323}	'a target'
Tone 4	high level	55	H	ba^{55}	'a dam'
Yin-ru	checked level	5	H	baʔ^{5}	'to shell'
Yang-ru	checked rising	35	MH	paʔ^{35}	'thin'
Neutral tone				ba	'(particle)'

'dipping' tone, which might be transcribed as [323]. (The corpus shows timing differences that make the two dipping tones more different from each other than these alphabetic transcriptions suggest, and these timing differences are perceptually salient, as shown by Fon 2000.) Thus, in the half century since Mandarin was enforced as the standard language of Taiwan, Guoyu has differentiated itself from Putonghua in ways that may eventually be as drastic as the differences between Putonghua and regional varieties of Mandarin within the People's Republic of China.

For the sake of comparison, Table 9.2 gives a traditional alphabetic representation for the Rugaohua citation-form tones. (Rugaohua is the regional variety of Rugao and neighbouring counties in Jiangsu Province, and classified as a subgroup of Mandarin, Jianghuai Mandarin, with about 67 million speakers—see Li, R. 1985, 1989*a*, *b*.) There are two 'extra' lines in the table, illustrating the two *rusheng* tones on 'checked' syllables. As in all varieties of Jianghuai Mandarin, Rugaohua still retains some vestige of the syllable-final stops of Middle Chinese, although the three-way place contrast of Middle Chinese has been lost, with historic /p/, /t/, /k/ all replaced by final glottal stop, as in the Wu varieties of Chinese. Disregarding the correspondences for these checked syllables, and comparing only the four types in syllables with all-sonorant rhymes, we can see that Tone 2 and Tone 3 are similar to the corresponding tones in Putonghua, but Tone 1 and Tone 4 are not.

9.2.3. *The 'neutral tone'*

The lexical tone system of Mandarin differs from Cantonese in that some morphemes, such as the agreement-soliciting particle -*ba*, the pragmatic particles -*ma* and -*a*, the verbal suffix -*le*, and the nominal suffix -*zi*, are inherently unspecified for tone. These morphemes are said to be in the 'neutral tone', as shown in Table 9.1. (Sometimes the lack of specification is identified even more explicitly as a tone type, by calling it 'tone 5'.) Containing a neutral tone syllable is one of the clearest indicators that a recurring sequence of syllables is a fully lexicalized polysyllabic word rather than a more decomposable compound or even a phrase, as in the monomorphemic *dōngxi* 'thing' versus the obviously compound *dōngxī* 'east–west'. However, Mandarin varieties differ in how reliable an indicator of lexical status this is—i.e., in the proportion of polysyllabic words that do contain neutral tone. In this respect, the old Beijing dialect that Chao was describing was almost a 'word accent' language when compared to modern Putonghua (Hu 1987). And Putonghua, in turn, makes greater use of neutral tone than does Guoyu (Tai 1978).

The phonetic value of the neutral tone is traditionally characterized as being 'parasitic' on (or predictable from) that of the preceding full-tone syllable. However, the exact nature of this 'parasitic' relationship seems to differ across varieties (and, possibly, across different phrasal contexts within some varieties). The difference between varieties is illustrated in Figures 9.1(a) and 9.1(b), which show fundamental frequency (Fo) traces for a Guoyu utterance and a Rugaohua utterance of a sentence containing the word *háizimen*. In the Rugaohua utterance, the high Fo value that is reached at the end of the Tone 2 of the initial syllable is maintained over the following two neutral-tone syllables. In the Guoyu utterance, by contrast, the Fo reaches a peak at the end of the *hái* and then falls gradually to a mid level at the end of that phrase.

In Putonghua as well as Guoyu, such a fall in pitch to a mid level is typical of citation form utterances of words ending in a neutral tone after Tone 1 or Tone 2. This is illustrated in Figures 9.2(a) and 9.2(b). When the neutral tone occurs after Tone 3, on the other hand, the fall-rise of the lexical tone is spread out over both syllables, as shown in Figure 9.2(c). And similarly after Tone 4, as shown in Figure 9.2(d), the fall of the full lexical tone is spread out over both syllables, so that Fo reaches a very low value at the end of the neutral tone syllable.

From the above, the Rugaohua pattern might be described in terms of classical 'tone spreading' whereby the last tone target is copied onto syllables that surface with no tone specification of their own (Huang 1999). However, the Guoyu and Putonghua case clearly is more complicated. The pattern in sequences with Tone 3 or Tone 4 might be captured with the kind of 'spreading' that has been described for Shanghai by Jin (1985) and others. However, this will not account for the pattern in sequences with Tone 1 or Tone 2. One possible analysis is that these citation form cases show something like the 'final lowering' posited for English by Liberman and Pierrehumbert (1984), among others. Perhaps there is even a L% boundary tone here (see Section 9.2.6). If this analysis is borne out, then we can posit

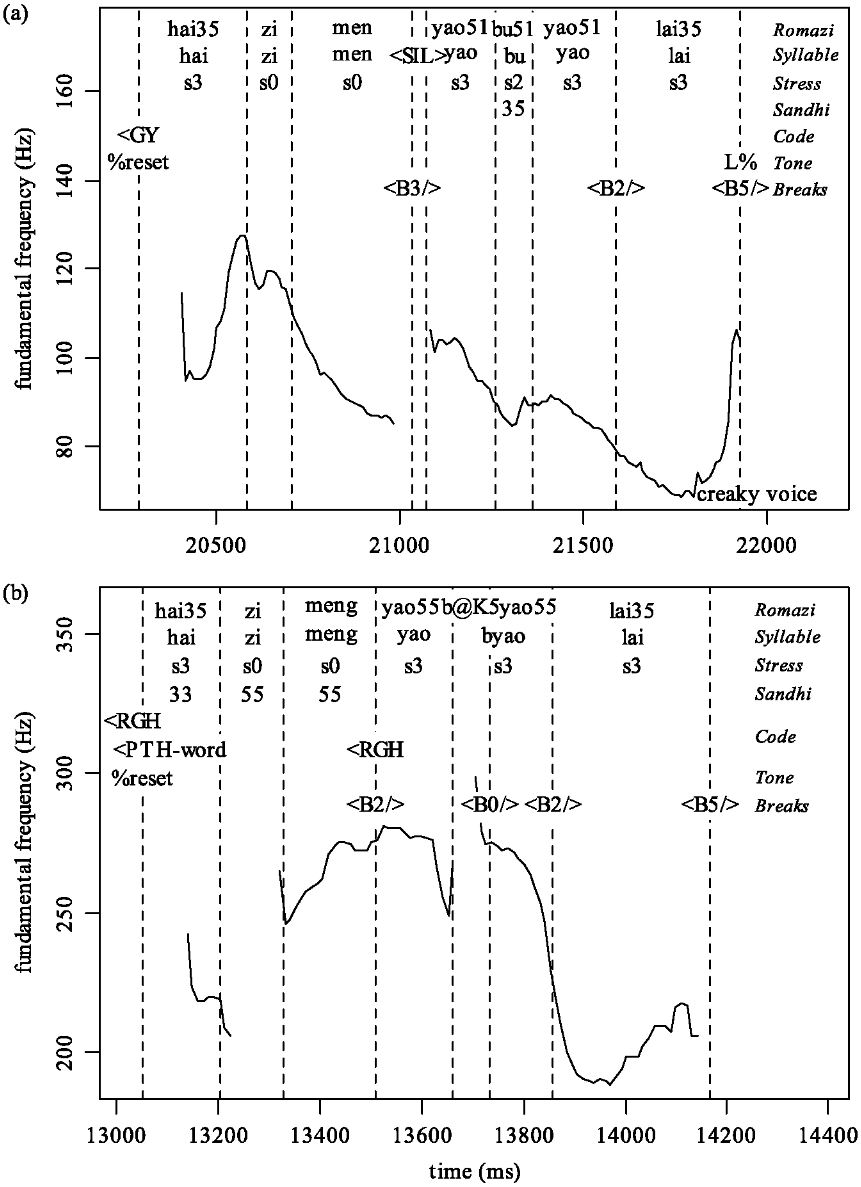


Figure 9.1 (a) Guoyu utterance of *Hái zi men yào bù yào lái?* 'Do the children want to come?' (b) Rugaohua utterance of same sentence as in Figure 9.1(a).

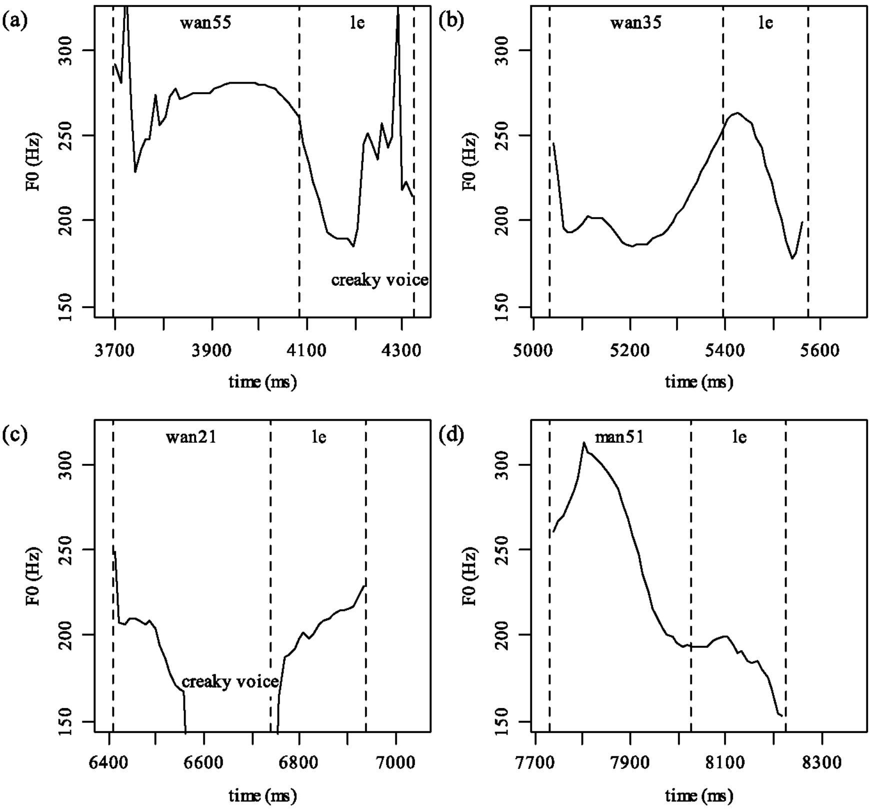


FIGURE 9.2 Putongua utterances of disyllabic predicates illustrating the realization of the neutral tone on the aspect particle *le* after each of the four full tones: (a) *wān le* 'to have (been) bent', [neutral tone after Tone 1], (b) *wán le* 'to have finished', [neutral tone after Tone 2], (c) *wǎn le* 'to have been late', [neutral tone after Tone 3], and (d) *màn le* 'to be too slow', [neutral tone after Tone 4].

a phrase edge with a L% boundary tone after the *háizimen* in Figure 9.1(a). Controlled observation of Guoyu and Putonghua neutral tone syllables in other phrasal contexts after each of the four lexical tones is needed.

In addition to having no independent tone specification, syllables with neutral tone are also characterized by segmental lenition. For example, they have shorter durations than syllables with a full lexical tone specification, and the inherently short high vowels can be devoiced after voiceless fricatives in these syllables when they follow a Tone 4 syllable, as in *dòufu* 'beancurd, tofu' and *yìsi* 'meaning' (Chao 1968: 37, 141). In neutral tone syllables in Guoyu and Putonghua, unaspirated stops are typically voiced, as if foot-internal, as in *zhīdao* 'to know' where the /t/ in the syllable, *dao*, is [d] phonetically. One also finds contractions, such as *tàm* 'they' alternating with the full, disyllabic form, *tāmen*, consisting of the third person singular form, *tā*, plus the *-men* plural suffix (Chao 1968: 141). This connection between segmental lenition and tone status is reminiscent of the accentual system of Swedish, where 'word accent' defines a categorical level of stress.

9.2.4. *Stress*

Thus, by contrast to Cantonese, Mandarin has stress. Moreover, it has it both inherent in the lexical entry for some morphemes (the neutral tone discussed above) and at the phrasal level. For example, the perfective aspect marker *-le* in *xiě-le* 'written' and the second syllable of *kuàizi* 'chopsticks', unlike the initial syllables, are inherently short and reduced and not specified for tone. The word *bù* 'not', on the other hand, is lexically a high-falling tone, but is produced as a weak neutral-toned syllable in the A-not-A interrogative construction (e.g., *liàn bu liàn* literally 'practise-not-practise' = 'Will (you) practise?'), unless the pragmatic context puts metalinguistic contrast on it.

Additionally, there is *local manipulation of pitch range* that might be called 'stress' at a higher level. This is illustrated in Figures 9.3 and 9.4. The sentence in Figure 9.3(a) was produced with narrow focus on the subject, *Wèi Lì*, in the utterance, *Wèi Lì mài làròu* 'Wei Li sells (Chinese) bacon', where the pitch range of the two focused syllables, *Wèi Lì*, is expanded, in contrast with the rest of the sentence. Such pitch range expansion does not occur in the corresponding sentence in Figure 9.3(b) with broad focus.

Jin's (1996) and Xu's (1999) data (for Putonghua speakers) suggest that these pitch range effects are not co-extensive with any phrasal unit. That is, the pitch range expansion is most obvious on the focused word, whereas the compression afterward extends over the entirety of the remainder of the phrase. Also, there can be 'pauses' in the middle, as in Figure 9.4(c) (for a Guoyu speaker).

Because the utterances in Figure 9.4 consist entirely of Tone 1 syllables, they also show clearly another aspect of these pitch range manipulations. When narrow focus is put on the first word in Figure 9.4(b), the proper noun *Ōuyīng*, the second syllable is affected more than the first, resulting in a rise over the word. Xu's (1999) data show a similar trend for the disyllabic noun *māomī* 'kitty' in initial position. It is tempting to relate this targeting of the second syllable to Chao's (1968) observation that sequences of full-toned syllables are not all equally prominent. That is, he describes a contrast

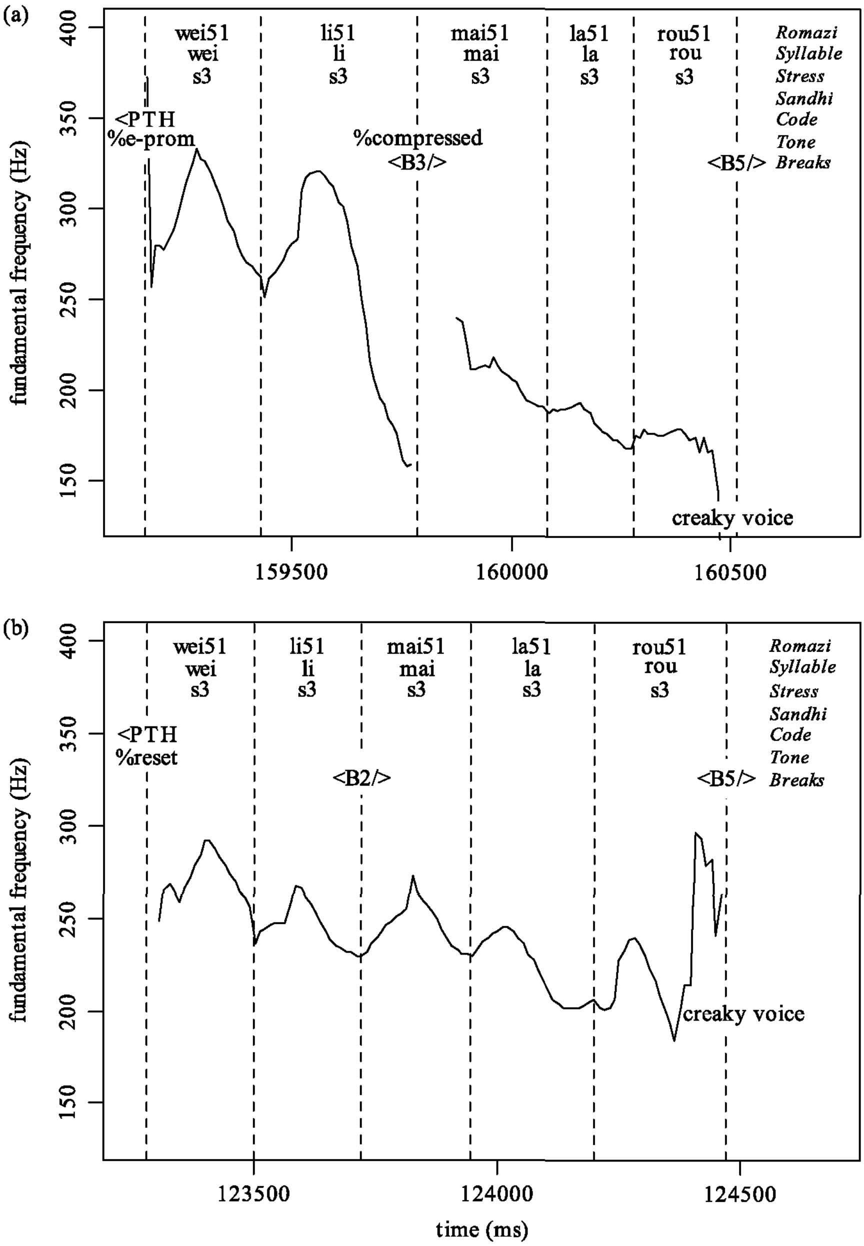


FIGURE 9.3 (a) Putonghua utterance of sentence *Wèi Lì mài là ròu*, 'Wei Li sells (Chinese) bacon', as a statement with narrow focus on *Wèi Lì*. (b) The same sentence uttered as an 'out of the blue' statement with broad focus.

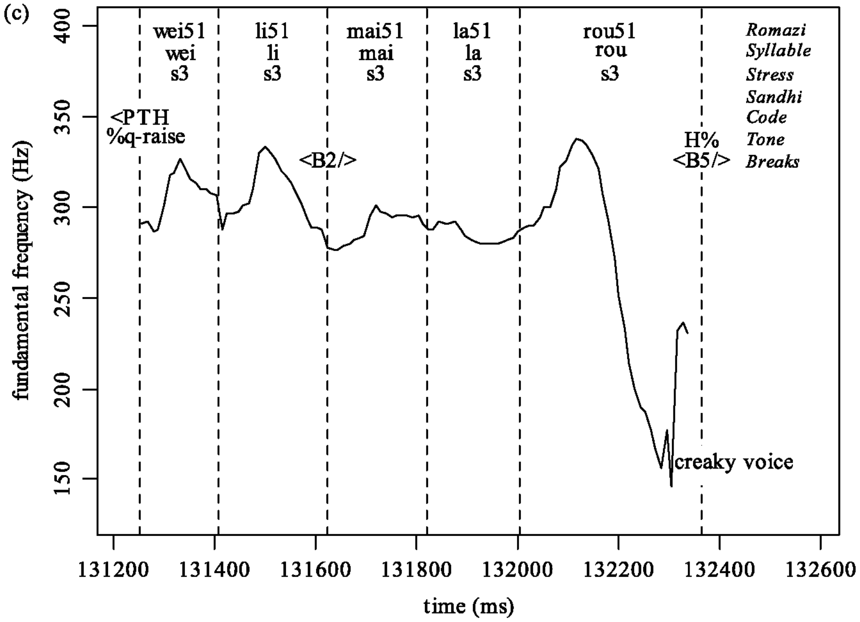


Figure 9.3 (c) The same speaker producing the sentence in Figure 9.3(a) as an echo question: 'Wei Li sells (Chinese) bacon!?'

between trochaic stress on forms such as *kuàizi*, but iambic stress on forms such as *māomī*, where there is stronger stress on the second syllable.

Chao describes this stronger stress on the last full-toned syllable of a word as consistent and predictable. It occurs for both phrases and compound words (Chao 1968: 360–1), for example. Thus, the stronger versus weaker stress in different positions are 'all allophones of one phonemic stress' (1968: 35), and the only phonemic contrast is between iambs (with two full-toned syllables) and trochees (with neutral tone on the second syllable). By contrast, Yin (1982) claims a three-way contrast among disyllabic forms with two full-toned syllables. In addition to Chao's 'predictable' iambic pattern of 'medium' stress followed by 'heavy' stress, he cites trochaic forms with 'heavy' stress on the first and 'medium' stress on the second syllable, and trochaic forms with a 'light' second syllable. Yin's own examples include minimal pairs for both of these types of trochaic forms. For the heavy-medium versus heavy-light cases, he gives *tèwù* 'special tasks/duties' versus *tèwu* 'secret agent, spy,' where the tone on the second syllable is frequently 'neutralized' in Putonghua. And for the medium-heavy (iambic) versus heavy-medium (trochaic) cases, Yin gives *sànbù* 'to take a walk' versus *sànbù* 'to scatter'.

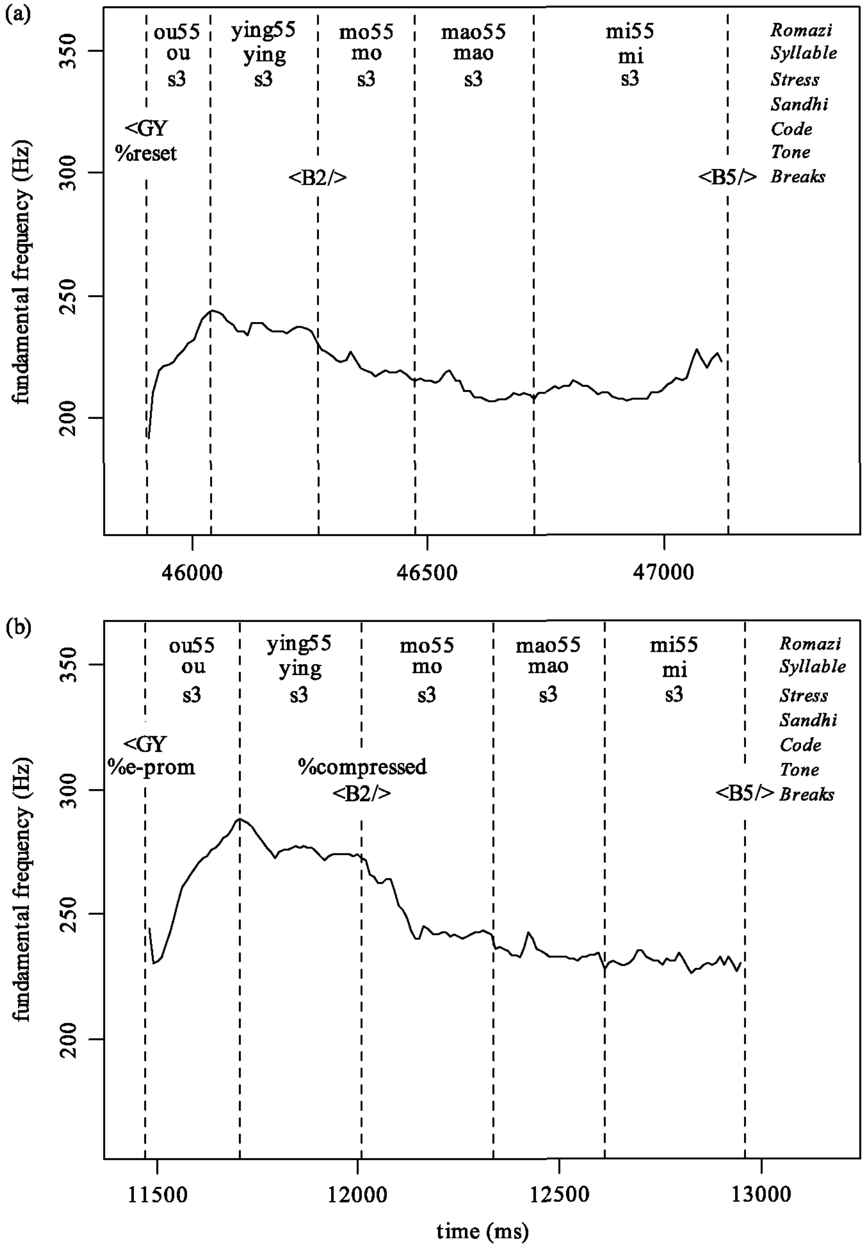


FIGURE 9.4 (a) Guoyu broad-focus utterance of *Ōu yīng mō māo mī*, 'Ouying strokes kitty'. (b) The same speaker's utterance of the sentence with narrow focus on *Ōu yīng*.

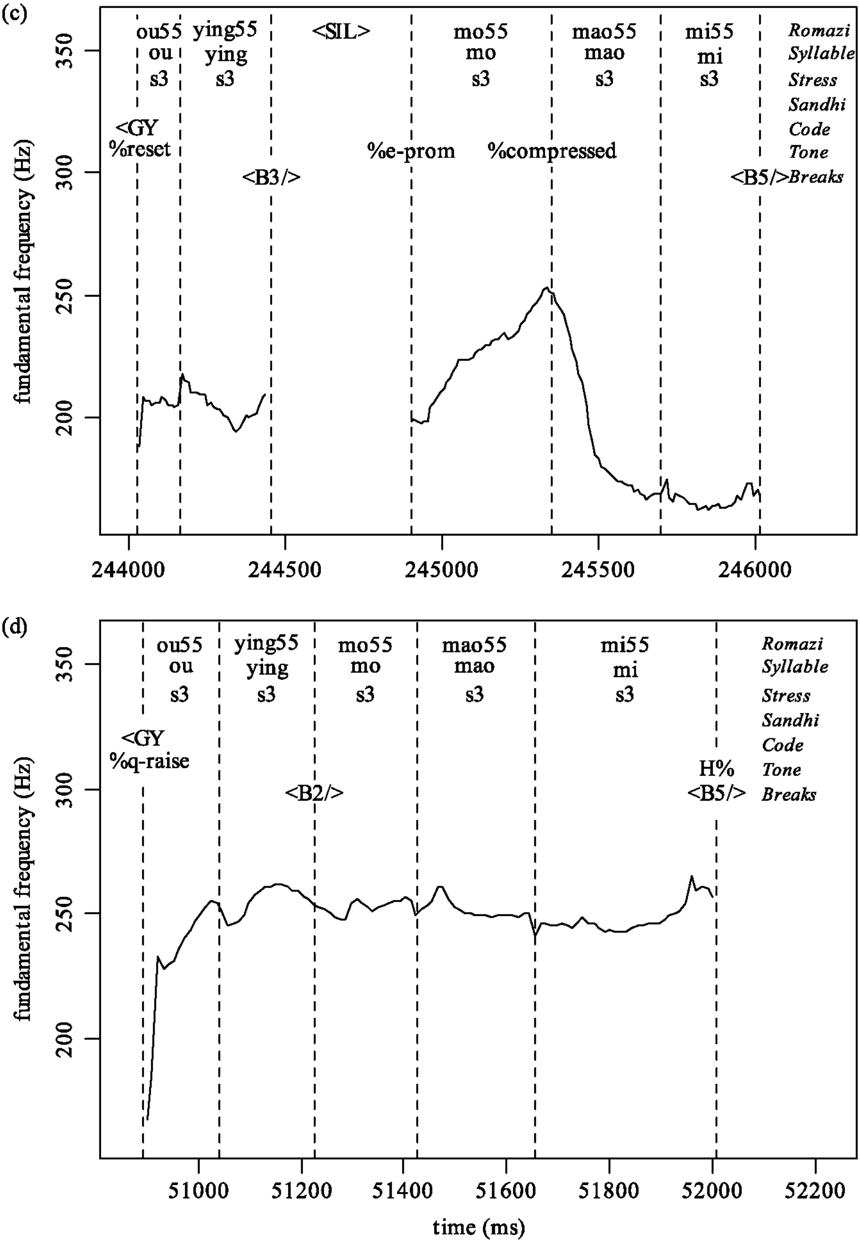


FIGURE 9.4 (c) The same speaker's utterance of the sentence with narrow focus on *mō*, and pause before it. (d) The same speaker producing the sentence in Figure 9.4(a) as an echo question.

Our attempts to elicit native speaker judgements do not support either Chao's claim or Yin's. That is, most native speakers of standard Mandarin deny any differences in stress level beyond the palpable contrast between (all) full-toned syllables and the neutral-tone syllables. However, stress is notoriously difficult to access with introspective judgements, and we need to look for evidence in patterns of production, such as potential contrasts in the domain of the pitch expansion under narrow focus, as in the particular targeting of the second syllable that is depicted in Figures 9.3 and 9.4.

Another place to look for evidence of a contrast between 'strong' (or 'primary') stress and 'medium' (or 'secondary') stress on full-toned syllables is in the possibility (and likelihood) of tone 'neutralization' in running speech, as in the *bu* of *liàn bu liàn* cited above. Just as varieties differ in the distribution of neutral tones in the lexicon, they also differ in regards to where neutral tone can occur in sentences. The A-not-A construction is a good illustration both of the effect and of the differences between Putonghua and Guoyu. The *liàn bu liàn* example cited above is Putonghua, and the second repetition of the verb is optionally also in neutral tone (Chao 1968: 270), making a ternary foot. The full form of the verb in a short declarative (as in the answer 'I'll practise') would be *liànxí*, or in connected speech *liànxi*, a more normal binary foot. A (rather pedantic) alternative to *liàn bu liàn* would be to repeat the whole verb: *liànxi bu liànxi*. In this case, the second *liàn* could not be in neutral tone. In Guoyu, the rhythmic constraints differ, and so does the syntax. The verb *liànxí* does not alternate readily between full and neutral tone on the second syllable, but is almost always produced with an identifiably rising Tone 2 on the second syllable (albeit somewhat reduced relative to the pitch fall on the first syllable). Also, the A-not-A construction incorporates the partial reduplication of the homologous Taiwanese construction so that the normal way to ask 'Are you going to practise?' is *liàn bu liànxí*, with neutral tone only on *bu*. That is, the second *liàn* is foot-initial and never reduced in Taiwan Mandarin.

As we have already suggested, the trochaic words with neutral tone can be described in terms of the stress foot. Neutral toned syllables cannot occur in isolation, and there are no content words that have only neutral-tone syllables. In many descriptions in the framework of Autosegmental Phonology and Metrical Phonology, this is accounted for by saying that a full-toned syllable can be a foot on its own, whereas a neutral-tone syllable is necessarily footed together with the preceding full-toned syllable, making the prosodic word in Mandarin minimally a bimoraic foot (see Yip 1980; Wright 1983; Duanmu 1990, *inter alia*). Shih (1986, 1997) proposes a somewhat different definition of the foot, in which even a full-toned monosyllabic form does not form a foot except under exceptional circumstances. (Beattie's (1985) definition of the 'prosodic word' is similar, as is Chen's (2000) 'minimal rhythmic unit'.) Shih's foot thus emphasizes the predominately disyllabic rhythm of the language, a metrical organization that is reflected in the lexical statistics (i.e. the much higher prevalence of disyllabic words in Mandarin than in Cantonese) as well as in some accounts of third tone sandhi, covered in the next section.

9.2.5. *Tone sandhi and the 'superfoot'*

Some of the lexically specified tones are subject to tone sandhi conditioned by tonal context and some other prosodic factors. The facts are clearest for Rugaohua. In this variety, the tone sandhi processes are similar to those found in the neighbouring Wu dialects such as Shanghainese (see Jin 1985; Selkirk and Shen 1990; Duanmu 1997; Chen 2000) in that only the first syllable in a sandhi domain keeps its original tonal specification while a rightward spreading forces the other syllables in the word or phrase to lose their underlying tones. Just as in Shanghainese, the sandhi domain in Rugaohua is constrained by such factors as syntactic structure, focus, speech rate, and so on. Some examples are shown in (1) below:

(1) (*a*) /da^{41} + ɕiaʔ^{35} + sə̃^{41}/ → [da^{41} ɕia(ʔ)^{11} sə̃^{41}] 'college student'
(*b*) /ɕiɔ^{323} + ɕiaʔ^{35} + sə̃^{41}/ → [ɕiɔ^{32} ɕia(ʔ)^{33} sə̃^{41}] 'elementary school pupil'

Each of the above three-syllable utterances has two sandhi groups, so that only the first and third syllables keep their underlying lexical tones. This corresponds to the left-branching syntactic structure shown in (2):

(2) [[X Y] Z]

The same three-morpheme sequences could be uttered with the first and second syllables bearing their lexical tones, and the third syllable grouped together with the second, as in (3), but then the syntax is right-branching, and the interpretation different.

(3) (*a*) /da^{41} + ɕiaʔ^{35} + sə̃^{41}/ → [da^{41} ɕia(ʔ)^{35} sə̃^{55}] 'big student'
(*b*) /ɕiɔ^{323} + ɕiaʔ^{35} + sə̃^{41}/ → [ɕiɔ^{323} ɕia(ʔ)^{35} sə̃^{55}] 'little student'

Stress patterns change accordingly. In a phrase with the sandhi grouping pattern of (1), syllables X and Z are stressed while Y is relatively unstressed, whereas in the pattern of (3), syllable Y is stressed and Z is relatively unstressed. Whether this is secondary stress or a total reduction to the level of a neutral tone syllable remains a question for further investigation.

The standard varieties of Mandarin also have tone sandhi. Most notably, Tone 3 changes to Tone 2 when followed by another Tone 3, and the occurrence of this 'third tone sandhi' seems to bear some relationship to stress, or at least to the disyllabic rhythm mentioned above. However, there is much that is still not understood about the phonetics of third tone sandhi, despite much discussion on the topic in the literature (e.g. Chao 1968; Cheng 1973; Shih 1986; Kratochvil 1987; Zhang 1988; Hung 1989; Shen 1990; Hsiao 1991; Chen 2000). Thus, Cheng (1973) emphasizes the blocking of third tone sandhi at certain syntactic (or prosodic) boundaries (reminiscent of the facts just described for Rugaohua), whereas Shih (1986, 1997), Hsiao (1991), and Chen (2000) point to rhythmic constraints, and posit a 'superfoot'—i.e. a grouping of a stranded monosyllable together with the adjacent disyllabic unit. (Shih (1997) reviews other prosodic factors as well, such as the 'refooting' of syllables that might otherwise be foot-final under conditions of prosodic prominence for narrow focus. And she points to 'morphological' factors such as sequence frequency.) Even the categorical nature of the sandhi change is disputed. For example, Wang and Li (1967) and Peng (1996) show that the sandhi tone is indistinguishable from underlying Tone 2 in identification tests, although Zee (1980) and Peng (1996) both show 'incomplete neutralization' in controlled instrumental studies. It is also possible that the facts differ across the major standard varieties.

Chao (1968) also describes various 'phonetic' tone sandhi changes. For example, Tone 2 in a trisyllabic sequence before a full-toned syllable changes into Tone 1 when the first syllable is Tone 1 or Tone 2. Underlying *cōngyóubǐng* 'onion oil cake', for example, is pronounced *cōngyōubǐng* in styles of Putonghua that are close to Beijinghua. There is also 'morphophonemic' tone sandhi for the word *bù* 'not' and the Tone 1 numerals *yī* 'one', *qī* 'seven', and *bā* 'eight'. For example, *bù* becomes Tone 2 before a Tone 4 morpheme—e.g., *bùgǒu* 'not lax' versus *búgòu* 'not enough'. These other sandhi phenomena and their relationship to stress and/or grouping are even less well understood than is third tone sandhi. Thus, much work is needed before we can say definitively whether or not tone sandhi phenomena define a level of grouping comparable to the clear sense of a 'phonological phrase' that the Rugaohua tone sandhi imparts.

9.2.6. *Evidence for higher-level prosodic grouping*

In addition to lexically specified tones, sometimes there are also pragmatic tones at the ends of sentences. Chao (1968: 812) posits a Rising Ending and a Falling Ending for Beijinghua, and analyses them as 'particles' on phrases and sentences. That is, he describes them as localized to the last syllable (specifically, to the voiced portion of the last syllable) of an utterance, revising an earlier analysis (Chao 1933) in which they are treated as part of a more global backdrop intonation contour. If Chao's observation is correct, his 'Ending' patterns would be tagged as boundary tones in a ToBI framework system.

Chao's observations were of Beijinghua from before the emergence of a separate standard in Taiwan. In our observations to date, we have identified at least two boundary tones also for Guoyu and Putonghua, which we tentatively identify with Chao's Rising Ending and Falling Ending (although unlike Chao, we have not observed them yet on syllables other than sentence-final particles, which are lexically unspecified for tone). One has a higher tonal target than the other. In Figure 9.5(a), *tāmen bú mài yúsǎn ma?* 'Don't they sell umbrellas?', the sentence-final particle *ma* is produced with a high boundary tone. The sentence in Figure 9.5(b), on the other hand, ends with a low boundary tone. In the first case, the speaker is asking a yes-no question, but the boundary tone suggests a presupposition that the store should sell umbrellas. Thus, this can convey surprise, if the addressee is someone who was sent to buy an umbrella and came back empty-handed. In the second case, by contrast, the L% boundary tone effectively makes the utterance a statement. It might be produced by a speaker to soften an explanation of why he came back empty-handed. The English equivalent might be something like, 'Well, but they don't sell umbrellas.'

In addition to these boundary tones, there are also global pitch range effects that can signal contrasting pragmatic meanings. For example, Figure 9.4(a) is a statement, *Ōuyīng mō māomī*, meaning 'Ouying strokes kitty'. When the utterance is produced as an echo question, the overall pitch range is raised, as shown in Figure 9.4(d). In older impressionistic descriptions such as Chao (1933), there was no differentiation between these global 'intonations' and the more localized boundary tone effects described above. As already noted, Chao (1968) implicitly recognizes the difference, but does so by describing only local effects, and Kratochvil's (1998) account is similar. However, some recent instrumental research, such as Shen (1990), has ignored the local effects to focus on the interaction of sentential pitch range effects with the lexical tone specification in short simple sentences such as the utterances in Figures 9.4(a) and 9.4(d). Longer utterances in more varied pragmatic contexts than those provided by Shen seem to show both global and local effects, interacting in ways that seem quite complicated, given our current limited understanding of the levels of prosodic grouping that are available to be the domain of phrasal pitch range specification. For example,

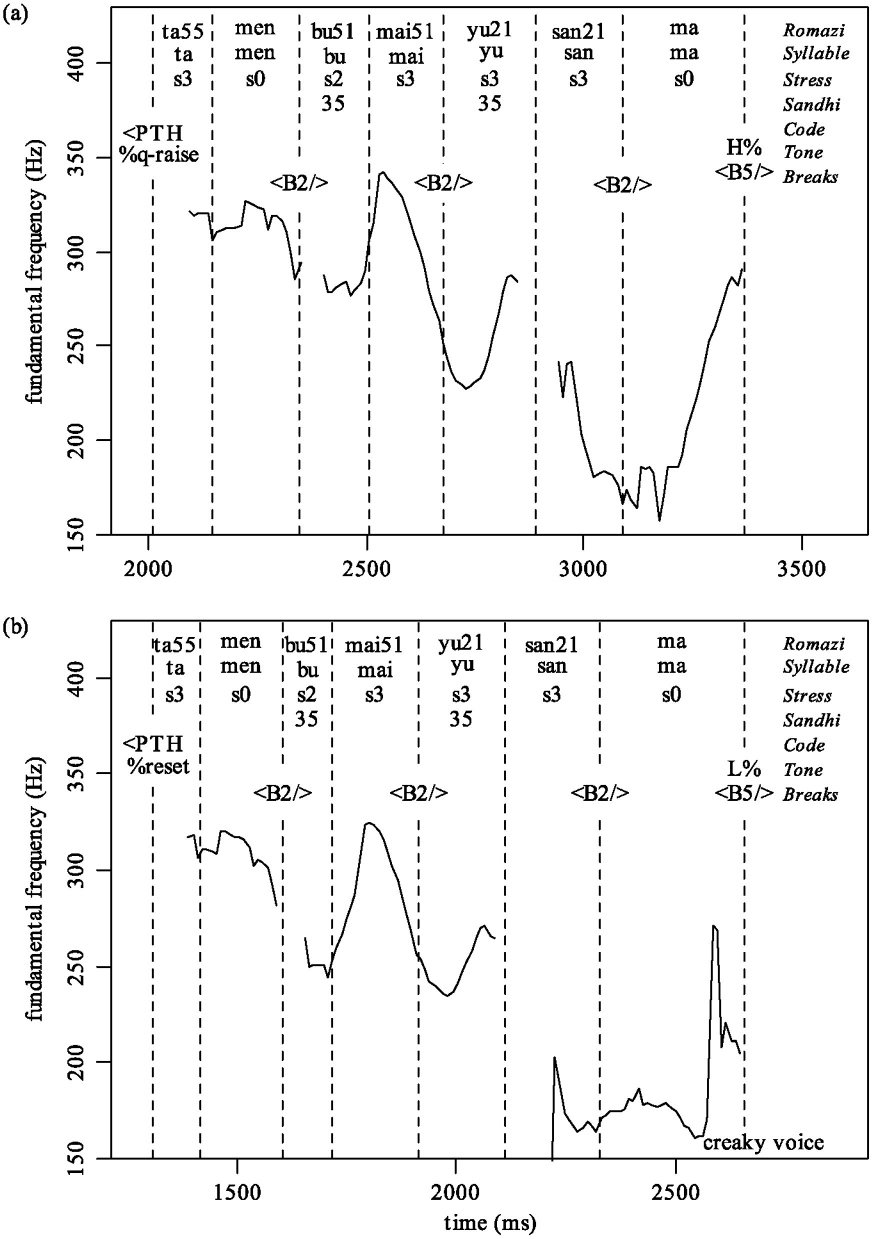


Figure 9.5 (a) Putonghua utterance of *Tā men bù mài yǔ sǎn ma?* 'They don't sell umbrellas?' produced with H%. (b) The same utterance as Figure 5(a) produced with L%.

some of our observations of Putonghua speakers suggest a three-way distinction among sentences ending with *ma*. There is a 'higher than neutral' ending, with a raising that does not start at the beginning of the sentence, but which also does not seem to be confined to the last syllable. And this pattern contrasts both with a 'neutral' final lowering and with a 'lower than neutral' ending (see Lee 2000 for a discussion of the phonetics and pragmatics of these effects). Substantial new studies are necessary to identify the domains of these effects, and to specify their relationship to the domain of the boundary tones.

These studies also should control carefully for factors such as prominence, which Jin (1996), Xu (1999), and others have also shown to affect pitch range (see Section 9.2.4). They also should investigate the relationship to the domain of tone sandhi in varieties such as Rugaohua (see Section 9.2.5). While any definitive statement is premature, it seems safe to speculate that the boundary tones and global pitch range effects can be identified with a larger prosodic grouping that is comparable to the 'intonation phrase' of Cantonese.

9.3. MANDARIN ToBI

Although our current understanding is not complete enough to propose a definitive set of annotation conventions for any variety of Mandarin, developing a preliminary set of tags for the phenomena described in Section 9.2 helps to clarify the outstanding questions for further research. A preliminary codification also provides a useful tool for investigating many of these phenomena using existing spoken language corpora. Since Mandarin is an economically and politically important language, it comes as no surprise to find that there are both proprietary and publicly available corpora. For example, the Linguistic Data Consortium has databases for both the PRC and Taiwan standard varieties. It is also not surprising that there have been several independent efforts to develop systems for tagging Mandarin corpora within the ToBI framework.

In this section, we will describe the union of two of these systems, one developed at Academia Sinica (AS) in Taipei and the other at the Ohio State University (OSU) in Columbus. In preparing for the satellite workshop at ICPhS'99, these two sites agreed to merge their systems into a single set of conventions that eventually might be applied to utterances, both spoken and read, in any variety of Mandarin—i.e. a Pan-Mandarin M_ToBI system. The two original systems were developed for rather different purposes. The AS conventions were developed for purposes such as developing the prosody component of a Guoyu text to speech (TTS) synthesis system, and training and testing automatic speech recognition (ASR) models using read speech. (It has also been adopted by the Institute of Linguistics at the Chinese Academy of Social Sciences in Beijing, to develop a Putonghua TTS system—see Li, A.-J. *et al.* 1999.) The OSU conventions were developed for purposes such as exploring the relationship between prosody and pragmatics using spoken language data, and documenting prosodic variables for sociolinguistic models of linguistically heterogeneous speech communities, including but not limited to Taiwan. Thus, the merging of the two systems is in keeping with the spirit of the original ToBI framework system for English. We are trying to forge a communal standard that can be shared across a viably large and varied community of users.

At this writing, the development of the merged system is not complete. Determining the final set of M_ToBI labels for Guoyu or Putonghua alone will require several more iterations of the same process that went into the development of the original AmEng_ToBI. (That is, participating sites contribute to a pooled set of calibration utterances that everyone transcribes. We calculate inter-transcriber agreement and discuss major discrepancies. We agree on a manageable set of changes to accommodate the problems that have emerged. The next iteration then tests the modified system against a new set of calibration utterances.) And codifying the full Pan-Mandarin system will be an even longer-term process, since our understanding of intonational phenomena in even the major regional varieties is very sketchy by comparison to research on either of these two standard varieties.

Even in its current unfinished state, however, the merging of the systems has already served to highlight several of the most salient prosodic phenomena of the language—particularly phenomena that were being treated similarly in the two original systems. Most noteworthy among these is the set of phenomena described above in Section 9.2.4. Both sites had independently chosen to tag syllable prominence using an explicit set of hierarchical labels (for 'emphasis' or 'stress' levels) that are separate from the tags for words, tones, and break indices common to all ToBI framework systems. A third system, developed at Lucent Bell Laboratories, is also for a Guoyu database, and it also tags stress levels. One of the first things that we hope to do in developing our Pan-Mandarin ToBI further is to conduct a three-site calibration experiment, in which each of us transcribes representative utterances from the other two sites, to develop mappings between these tags across all three systems.

With the caveat, then, that many details almost certainly will have changed in the final standard, we describe in this section the symbolic tiers of the current state of the merged M_ToBI system. We will relate these tiers to the

Table 9.3 The eight tiers

	Tier	Description
1.	Words	syllable-by-syllable transcription in Chinese characters
2.	Romanization	syllable-by-syllable transcription in modified Pinyin romanization using ASCII (hence, without tone diacritics)
3.	Syllables*	phonological syllables that do not correspond in a one-to-one relationship with orthographic syllables (e.g., for contractions as such as *tām* for the disyllabic pronoun, *tāmen* 'they').
4.	Stress	relative degree of stress marked on each syllable
5.	Sandhi	marking of tone sandhi (e.g., 35 for the third tone sandhi form of the first syllable in /yu21san21/), allotones (e.g., 214 for Tone 3 in an utterance final position), morpheme-specific tone sandhi (e.g., *bu35* 'not' in *bú yào* 'don't want', from *bù* + *yào*)
6.	Tones	marking of boundary tones and pitch range effects (e.g., %compressed at the onset of pitch range reduction)
7.	Break Indices	hierarchy of disjunctures to represent prosodic phrasing
8.	Code	identification of the variety of Mandarin and marking of points of code-switching

* In the Emu implementation of the M_ToBI standard, the stress tier and sandhi tier are simply two extra labelling fields for the syllables level, whereas in the xwaves/xlabel implementation, the relationship among these tiers must be insured by an independent grammar checker.

phenomena described above in Section 9.2. We also will identify other ways in which the two systems either were in accord or complemented each other enough to make it possible to conflate the two sets of tagging conventions. Eight tiers are proposed at this stage in the system. These tiers are listed in Table 9.3, and discussed in turn in the remainder of this section: the words and romanization tiers in Section 9.3.1, the syllables, stress, and sandhi tiers in Section 9.3.2, the tones tier in Section 9.3.3, the break indices tier in Section 9.3.4, and the code tier in Section 9.3.5.

9.3.1. *Transcribing the 'words'*

As in other ToBI framework systems, the symbolic tags in an M_ToBI transcription are intended to be anchored in time to an audio recording of the utterance. In ToBI framework systems for most other languages, the initial set of time stamps is via an obligatory *words* tier. The words tier provides a label for each word in the utterance in the native orthography (or in a romanized transliteration, in the case of Gr_ToBI—see Arvaniti and Balthazani, this volume Ch. 4). It thus orients the transcriber to the signal and accompanying F0 contour in terms of a familiar representation that is accessible to anyone who knows the language. It also promotes the development of tools for automating aspects of the transcription using resources such as online pronunciation dictionaries.

In keeping with these functions, the words tier for Mandarin is defined as a syllable-by-syllable transcription in Chinese characters. Although Mandarin has far fewer monosyllabic words than Cantonese, having a separate tag for each syllable accords with native speaker intuitions about what 'words' are in Chinese. Moreover, the use of Chinese characters provides a common set of easily accessible tags that 'normalize' over the substantial segmental differences among regional varieties. The OSU group has implemented M_ToBI in Emu (Cassidy and Harrington 2001), a database labelling/querying language that allows Chinese character input on a PC/Windows platform using standard encodings such as Unicode and Big5. However, the labelling platform that is common to both original sites (the UNIX/Linux based xwaves/xlabel program) does not allow Chinese character input. Therefore, M_ToBI currently does not specify the words tier as obligatory. That is, we allow the orthographic transcription to be stored in a separate text file, just so long as it can be linked indirectly to the audio file via some other tier that also provides a syllable-by-syllable transcription of the utterance.

In the current M_ToBI system, this other tier is by default the *romanization* tier (abbreviated as romazi in the label file extension). The romanization tier tags each orthographic 'syllable' using an 'ASCIIfied' version of the Pinyin system. This modified Pinyin romanization is strictly ASCII in that lexical tones are marked using Chao's tone numbers rather than with the original Pinyin tone diacritics. For example, Figures 9.5(a) and 9.5(b) show the lexical tone on all syllables except the plural marker, *men*, and the sentence-final particle, *ma*, both of which are in the neutral tone. A further modification from Pinyin is that each syllable is labelled with the 'dictionary form' for the particular variety in which the utterance is produced. That is, we have decided not to 'normalize' to a Putonghua reading for the Chinese character that would be used to transcribe the syllable in the associated text file (or the currently optional words tier). Some of the ramifications of this decision are illustrated in Figure 9.6.

The Pinyin romanization system was originally developed for Putonghua. We know that it is adequate for transcribing the other two national standards, Guoyu and Huayu, because they are segmentally very close to Putonghua. Pinyin also provides some 'extra' characters to transcribe sounds found only

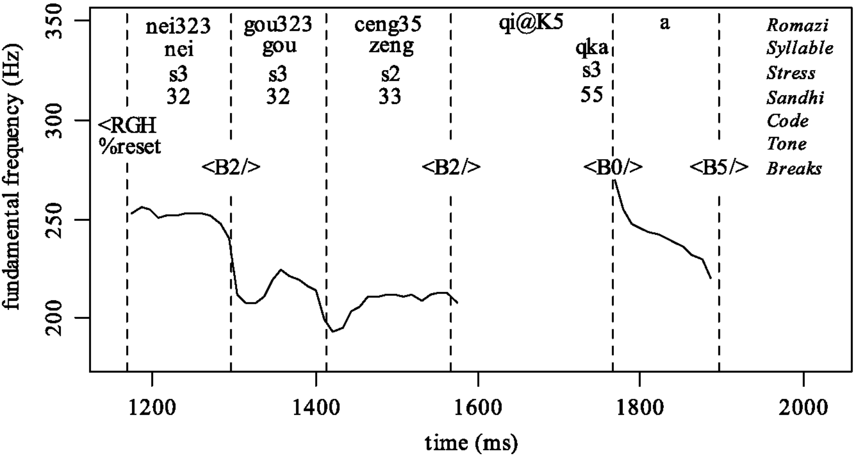


Figure 9.6 Rugao utterance of the sentence *Nei323 gou323 ceng35 qi@K5 a*. 'Have you eaten?' with tone spreading from *gou323* to *ceng35*, and from *qi@K5* to *a*.

in the prestigious Beijing regional variety. (Standard broadcast Putonghua is sometimes equated with Beijing Mandarin, but the two are distinct—see for example, Hu 1987; Chan and Tai 1989.) Other regional varieties require further additions, such as the 'K' that we have added to transcribe the final glottal stop in the Rugaohua word, *qi@K*^{5} (/tɕiəʔ^5/) 'to eat' in Figure 9.6. Before we can codify the romanization tier in the final M_ToBI system, we need to test how well Pinyin does with other regional varieties. If too many added characters are needed to cover the full range of segmental variation that is observed, it may be more tractable to replace Pinyin with something closer to a phonetic transcription. In that case, we will adopt instead the SAMPA-like transcription that the Academia Sinica group developed for segmental transcription.

While it is not part of the original ToBI framework system for prosodic transcription, the Academia Sinica (AS) Guoyu databases are all labelled with this phonemic alphabetic transcription of 'initial' (onset) and 'final' (rhyme) segments (see Tseng and Chou 1999). The transcription standard was designed to be compatible also with segmental transcriptions for Taiwanese and Hakka (Kejia), and it should extend easily to regional varieties of Mandarin, too. For standard Mandarin utterances that are transcribed at Academia Sinica, these segmental tags can be aligned automatically with the signal from the orthographic transcription file using the text preprocessing component of the site's TTS system in combination with an HMM recognizer (see Chou *et al.* 1998).

9.3.2. *Transcribing syllable prominence and tone sandhi*

As noted above in Section 9.2.3, orthographic syllables do not always correspond to phonological syllables, because of contractions such as *tām* for *tāmen* 'they'. The syllables tier allows us to capture such discrepancies, by labelling only the phonological syllables. In its current form, these syllables are transcribed in a broad phonetic transcription using the same Pinyin-based symbols as in the romanization tier. As more spontaneous speech is recorded and transcribed in the M_ToBI system, we may find it useful to differentiate this tier even more from the orthographic tier, by recording major allophones.

In the Emu implementation of the M_ToBI standard, the syllables tier (shortened to 'Syllable' in the figures here) has two other labelling fields that mark the relative degree of stress of each syllable in the utterance (stress tier) and any tone sandhi (sandhi tier). In the xwaves/xlabel implementation, the relationship among these three tiers must be insured by an independent grammar checker. We describe the stress and sandhi tiers in turn below.

The general idea of transcribing stress levels was common to both original systems. The two systems agreed even on the number of distinct levels to transcribe. In the AS system, the four labels E0 through E3 were defined in terms of 'reduced' versus 'normal' versus 'moderate' versus 'strong' levels of 'emphasis' for each syllable. However, by comparison to the development of the break index definitions (see Section 9.3.4), little attention had been given to examining how these definitions were understood by different transcribers. Thus, we need a thorough inter-site calibration experiment before we can be completely sure how the AS E0 through E3 map onto S0 through S3 at the other site. In the interim, the preliminary M_ToBI standard has adopted the more specific definitions of the four 'stress' levels in the OSU system. These levels are summarized in Table 9.4.

Stressed syllables with fully realized tone are labelled with *S3*. This level is illustrated in each of the utterances in Figure 9.5 by the first syllable of the

Table 9.4 Stress levels in the stress tier

S3	syllable with fully-realized lexical tone
S2	syllable with substantial tone reduction (e.g., undershooting of tonal target with duration reduction)
S1	syllable that has lost its lexical tonal specification (e.g., in a weakly-stressed position)
S0	syllable with lexical neutral tone (i.e., such a syllable is inherently unstressed)

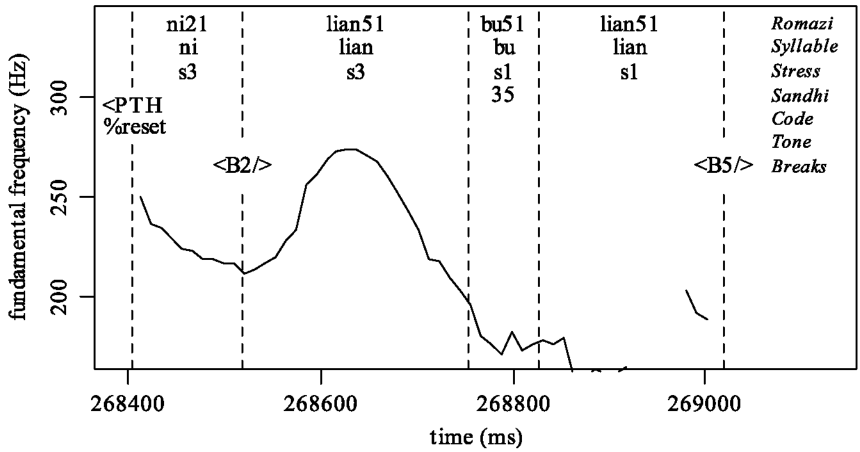


FIGURE 9.7 Putonghua utterance of *Nǐ liàn bù liàn?* 'Will you be practising?

subject *nǐmen* 'you (plural)' and all three syllables of the verb phrase *mài yúsǎn* 'sell umbrellas'. Syllables with lexical neutral tone are inherently unstressed and, therefore, are labelled with *So*. This level is illustrated twice in each of the utterances in Figure 9.5, where both the plural marker *-men* of *nǐmen* 'you (plural)' and the sentence-final particle *ma* are labelled with *So*. In running speech, some syllables with lexical full tones are often 'neutralized'. That is, they are produced with the tonal and temporal characteristics of lexical neutral tones. In Figure 9.7, in the A-not-A construction *liàn bú liàn*, the last two syllables, *bú liàn*, were neutralized. Both syllables have 'lost' their tonal specification. That is, there are no traces of a high-rising tone on *bú* or of a high-falling tone on *liàn*. Lacking definitive evidence that this tone 'loss' makes such syllables 'feel' as reduced as an 'inherently' neutral tone So syllable, these syllables are labelled with *S1*. *S2* is then used to label syllables with substantial tonal reduction that stops short of this complete 'neutralization' of tonal specification. That is, S2 labels a substantial undershoot of the tonal target, which is usually accompanied by at least some shortening of duration. For instance, in Figure 9.8 the high-falling tone *kuài* does not reach the very low tonal target suggested by the '1' in the tone transcription.

Note that we have characterized this fall to mid level in terms of a phonetically gradient undershoot of the tone target that is tagged only implicitly by the S2 label. By contrast, Chao (1948: 26) described it in terms of a tone sandhi alternation between a 'full' falling tone /51/ and a 'half' falling tone /53/. Later, Chao (1968: 28–9) accounted for the 'full' versus 'half' falling tone as the effect of stress, with greater stress enlarging both the range and the length of the

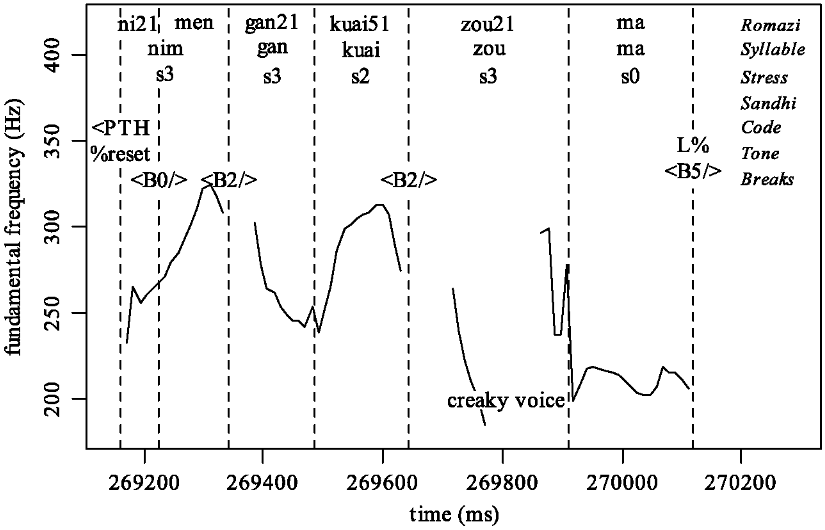


Figure 9.8 Putonghua utterance of the sentence *Nǐ men gǎn kuǎi zǒu ma*! 'You (pl.) should leave right away!'

tone. And since he analyses disyllabic words with full-tone on both syllables as having iambic stress, it is the first syllable that is realized with the partial falling tone. In this later formulation, Chao considered such stress-related phonetic differences to hold true for any combination of tones, and did not limit it to Tone 4. Thus, in both formulations (albeit more clearly in the later one), Chao distinguished this kind of 'phonetic' sandhi from the clearly categorical alternation involved in the 'third tone sandhi' that Taiwanese schoolchildren are taught in elementary school. He described the undershoot in terms of the alphabetic alternation only because he did not have access to a more gradient phonetic representation such as the Fo contour that is a necessary part of any ToBI framework transcription.

In the current M_ToBI system, then, we adopt Chao's distinction between 'phonetic' and 'phonological' sandhi. We use S2 alone to tag the variation in tone shape that arises from undershoot and coarticulation, but provide an explicit alphabetic transcription for variation such as third tone sandhi where there is clearer evidence that a categorical alternation is involved. (This follows the practice of the OSU system. The AS system circumvented the issue by not transcribing tone at all. Instead, the TTS prosody component was trained on the AS read-speech database to extract a 'tone pattern' that could be generated for each different sequence of tones that can occur within a minor prosodic phrase—see Section 9.3.4.) For categorical alternations,

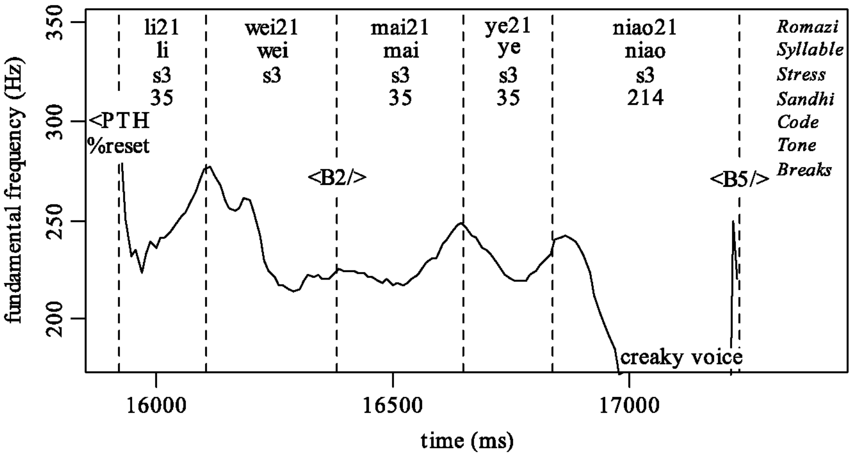


Figure 9.9 Putonghua utterance of *Lĭ Wĕi măi yĕ niăo*, 'Li Wei buys a wild bird.'

then, the sandhi form of the affected tone is marked on the sandhi tier with transcription in Chao's tone numbers. Thus, *35* represents the third tone sandhi of the standard varieties, where Tone 3 is phonetically realized as a rising tone. This is illustrated in Figure 9.9, where the first, the third, and the fourth syllables of the sentence were produced with the *35* sandhi tone. (Figure 9.6 shows analogous examples for sandhi forms in Rugaohua.)

The tone sandhi tier is also used to label two other types of alternation that are not always termed 'sandhi' in the literature. For example, *214* is used to label Tone 3 produced as a low dipping tone in any of the standard varieties. As noted above in Section 9.2.2, this is an 'allotone' of Tone 3 which may occur in sentence- or phrase-final position, or when the syllable is produced with emphatic prominence. It is illustrated by the last syllable in Figure 9.9. We can think of this alternation as position-specific tone sandhi. The other type of alternation is morpheme-specific tone sandhi. For example, the negative morpheme *bù* in the standard varieties has an alternate rising form when it is followed by another high-falling tone, as in *bú yào* 'don't want'. Such morpheme-specific variation is also transcribed on the sandhi tier, as shown by the rising alternate of *bù*, marked with *35* in Figure 9.7.

Note, however, that there are alternative analyses for both cases. For example, one traditional description of the /21/ alternate of Tone 3 is that it is a phonetic 'truncation' of a more canonical /214/ in prosodic positions where syllable duration is short for other reasons, such as not being phrase-final (although see Section 9.2.2 for difficulties with this analysis for Guoyu). In a similar way, since the negative morpheme *bù* 'not' is typically subject to 'neutralization' in running speech, it is possible to think of the *bú* alternate as a lexicalization of the 'neutralized' variant in an unusual 'upbeat' position, where the tone shape cannot be 'parasitic' to the preceding tone. That is, the rise might be described as a 'phonetic interpolation' to the high target at the beginning of the *yào*. If the 'upbeat' analysis is correct, then the 35 tagging on the sandhi tier for the rising alternate of *bù* should be replaced by an S1 specification on the stress tier. If the categorical sandhi analysis is correct, on the other hand, then the 35 tagging of *bù* on the sandhi tier should always be accompanied by stress level S2 or higher on the stress tier. In the current M_ToBI system, we do not enforce the sandhi analysis, but instead allow transcribers to combine *S1* with the 35 tag for *bù*.

As mentioned above in Sections 9.2.4–9.2.5, the relationship between stress and tone sandhi is a very under-researched area. It is also likely that the relationship is different for different varieties, and the conflicting analyses of the rising variant here may reflect real differences between the native varieties of the linguists proposing them. Explicitly tagging the alternation on an independent sandhi tier without requiring the stress level to be higher than S1 prompts us to explore these possibilities in a way that we could not do if we codified the system prematurely. Moreover, the tone sandhi patterns for different regional varieties are as varied as the segments and lexical tones. Having a sandhi tier independent of the more canonical tones tier (described in the next section) encourages researchers to develop labels that are specific to each new variety that is transcribed, without making a premature commitment to the relationship between any sandhi pattern and the intonational phrasing.

9.3.3. *Transcribing the 'intonation'*

Because the distinction between tonal reduction and tone sandhi is still an open issue, and because the evidence for boundary tones is less clear than in Cantonese, the current version of M_ToBI does not follow the C_ToBI convention of tagging lexical tones and boundary tones together on the same tier. Instead, we transcribe the 'underlying' lexical tone on the romanization tier, the potentially more rhythm related effects of tone sandhi on the sandhi tier, and make a third tier for the boundary tones and pitch range effects discussed in Section 9.2.6. That is, the M_ToBI version of the usual ToBI *tones* tier includes only these 'true' intonational effects.

The tones tier marks global or local characteristics of the backdrop pitch range, such as a general downtrend versus raised pitch for a sentence, as well as the phenomena that were described in terms of boundary tones in

Section 9.2.6. (The AS system circumvented the issue of whether these 'true edge events' can be distinguished from more global trends by training the TTS system to generate a stylized 'intonation contour' for each breath group, based on the punctuation of the read text—see Chou *et al.* 1996. In order to generalize to spontaneous speech and to more different read-speech styles, we included the 'intonation' tags of the OSU system in the merged M_ToBI standard. This is an example of a way in which the two systems complemented each other.)

Tags for boundary tones are like those in other ToBI framework systems. Thus, as shown in the set of tags for the tones tier in Table 9.5, *H%* and *L%* mark high boundary tone and low boundary tone, respectively. (See Figures 9.5(a) and 9.5(b) (where the tones tier is given as 'Tone'), and the discussion above in Section 9.2.6.)

Tags for backdrop pitch range effects are descriptive terms set off with the same '%' that identifies the boundary tones. For Putonghua and Guoyu, there are two types of pitch range effects that are tagged in this way: ones that describe a backdrop 'contour' for an entire sentence or phrase (see for example Shen 1989, for Putonghua), and ones that are associated with more localized effects of focal prominence (see for example Jin 1996; Xu 1999). However, the syntax is the same. The tag is placed at the beginning of the effect, which is also the beginning of the relevant phrasal unit in the first type of tag. Thus, *%reset* marks the beginning of a new pitch downtrend over a 'normal' declarative phrase (i.e., a pitch reset), whereas *%q-raise* marks the flat raised pitch range regularly seen in echo questions. As shown in Figure 9.3(b), the %reset label is placed at the beginning of the phrase, to signify the beginning of the pitch downtrend. Figure 9.3(c) shows the corresponding syntactically unmarked echo question, and the sentence-initial %q-raise tag. (Another pair of examples is the statement in Figure 9.4(a) and the corresponding echo question in Figure 9.4(d).)

TABLE 9.5 Tags for boundary tones and backdrop pitch range effects in the tones tier

H%	high boundary tone (at the end of an utterance)
L%	low boundary tone (at the end of an utterance)
%reset	beginning of a new pitch downtrend or pitch reset
%q-raise	beginning of a raised pitch range (e.g., in echo questions)
%e-prom	beginning of local expansion of pitch range due to emphatic prominence
%compressed	beginning of reduction of pitch range of syllables following the expansion of pitch range under %e-prom

The placement of the second type of pitch-range tag is illustrated in Figure 9.3(a). The beginning of a local expansion of pitch range due to emphatic prominence is marked with the tag *%e-prom*. The expansion is often accompanied by reduction of the pitch range of the following syllables, which is signalled by *%compressed* at the onset of pitch range reduction. The sentence in Figure 9.3(a) was produced with narrow focus on the subject noun, *Wèi Lì*. The beginning of the expanded pitch range over these two emphasized syllables is marked with %e-prom, and the beginning of the compression of pitch range on the succeeding syllables is marked with %compressed.

The tags described here reflect our current understanding of the standard varieties of Mandarin. The tones tier conventions for Rugaohua probably will have to be different. For example, our casual observation to date suggests that tags such as %e-prom can be linked to break indices for phrasal units at the level of the tone sandhi group, as in the neighbouring Wu varieties (see for example Selkirk and Shen 1990, and the discussion in Section 9.2.5). However, there is as yet no extended study to confirm this. Moreover, our knowledge of such intonational phenomena in other regional varieties is even sketchier. We should not assume that tones tier labels that are appropriate for Putonghua can be applied to any regional variety except as an initial working hypothesis in the development of variety-specific conventions.

9.3.4. *The break indices*

As in other ToBI framework systems, prosodic phrasing is represented with a hierarchy of break indices, with a number (a particular level of disjuncture) tagged at the end of every segment on the words tier and/or romanization tier. The merged M_ToBI system adopts the break-index conventions of the AS system which have been tested against the entire Guoyu TTS database, in an inter-transcriber calibration experiment involving the two primary transcribers (see Tseng and Chou 1999). The conventions are summarized in Table 9.6.

TABLE 9.6 Break indices and corresponding events on other tiers

B0	reduced syllable boundary	i.e. contraction; requires S0 or S1 on left or right
B1	normal syllable boundary	the 'default' case within a polysyllabic word
B2	minor phrase boundary	must be followed by at least S2
B3	major phrase boundary	
B4	breath group boundary	H%, L%, %reset, and %q-raise on the tones tier
B5	prosodic group boundary	

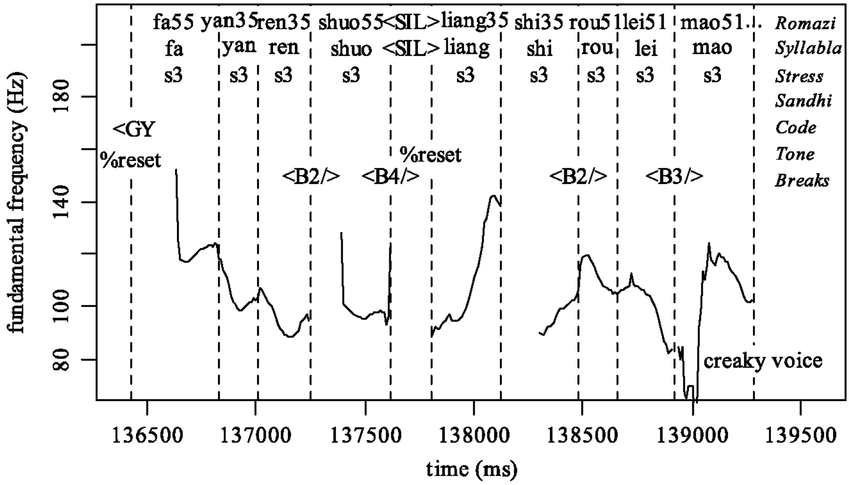


FIGURE 9.10 Fragment of Guoyu utterance *Fā yán rén shuō: Liáng shí ròu lèi mào yì yáng háng jiāng yú míng rì zhèng shì kāi yè*, 'The spokesperson said: "Farmers and Butchers Trading Company will open for business tomorrow."'

Six levels of prosodic juncture are distinguished. Level 4 marks boundaries which correspond to a reset of pitch between sentences or phrases. If a breath group boundary is accompanied by a prolonged pause, it is labelled with break index 5. The first of these two labels (breath group with only a small pause) is illustrated by the utterance fragment in Figure 9.10, which is part of a long sentence produced as two breath groups. There is a large reset accompanied only by a short pause after the verb, *shuō* 'to say', which sets off the quoted material from the matrix clause. In the AS calibration experiment, a large percentage of inter-transcriber disagreements involved confusion between B4 and B5. Even after exchanging notes and retranscribing, 21 percent of the breaks that were labelled with B4 or B5 by either transcriber were labelled with the other of these two tags by the other transcriber. This is unsurprising for a distinction which is not a categorical one, but a matter of gradient differences in the degree of final lowering before the boundary, the amount of pitch range reset after the boundary, and so on. In the final M_ToBI system, these two labels may be collapsed and the distinctions captured instead in a direct metric of pitch range that can be correlated with a recursive model of discourse structure. (Cf. also the 'finality tier' in the J_ToBI conventions described by Venditti, this volume Ch. 7.)

Break indices 2 and 3 are used for phrase boundaries within breath groups. Here, the M_ToBI conventions distinguish between major phrases and minor phrases. The two transcribers in the AS calibration experiment were instructed to transcribe B3 when they perceived a pause, and B2 when no pause was perceived. These two labels were next most confusable after B4 versus B5. Even after comparing notes and retranscribing, there was 9 percent confusion between these two tags. Figure 9.10 shows two examples of break index 2 and one example of break index 3, as identified by transcribers at the OSU site who were given the same instructions as the two transcribers in the AS calibration experiment. It is possible that further inter-site calibration tests will uncover better criteria for defining major versus minor phrases. Even if no categorical markers are identified for the standard varieties, however, these two levels of grouping probably should not be collapsed, since B2 probably will be needed for the tone sandhi group in Rugaohua.

Break index *1* is the 'default' boundary between syllables, and is left unmarked in the figures, to avoid clutter. (In transcribing with the xwaves/xlabel platform, these can be inserted automatically and then changed for other break index levels.)

Break index *0* tags boundaries that have been 'deleted' when a syllable is extremely reduced (S0 or S1) and its vestige segments re-syllabified with the previous syllable or the following syllable. For example, in Figure 9.8, the second syllable of *nǐmen* 'you (plural)' is reduced to just the initial bilabial nasal and re-syllabified with *nǐ*, so the boundary between *nǐ* and the plural suffix, *-men*, is labelled with B0.

9.3.5. *The code tier*

As noted above, the M_ToBI conventions are a 'Pan-Mandarin' system; they are intended to be applicable to utterances in any variety of Mandarin. The conventions also are intended to serve as a tool for exploring prosodic variables in complex speech communities where speakers typically are fluent in several varieties of Mandarin (or in Mandarin and a non-Mandarin variety of Chinese), and can switch easily between codes for sociolinguistic or stylistic effect. We have noted at several places in the discussion above that the current inventory of labels may need to be expanded or modified to cover other varieties besides Guoyu and Putonghua. Rather than 'recompile' the inventory of labels each time M_ToBI is extended to a new regional variety, it would be convenient to add labels or provide variety-specific interpretations of old labels, as appropriate for the system. This more modular approach, however, requires that we identify the variety being transcribed so that labels that are similar across varieties can be interpreted accordingly. (See Bruce, this volume Ch. 15 for a discussion of regional variation in the phonetics of the accent 1 versus accent 2 contrast in Swedish. See also Grice *et al.*, this volume Ch. 13, for the more extensive variation that needs to be covered in developing a Pan Italian ToBI.) Ideally, the mechanism that identifies the variety can then also serve to identify points where the speaker incorporates features of another variety in prosodic code-switching.

The *code* tier is being developed for these two related purposes. Labels on this tier identify the variety of Mandarin used to produce the utterance and also mark points of code-switching between varieties. The syntax of these labels is similar to that of tags on the miscellaneous tier in English. Each tag consists of a start marker '< ' followed by an abbreviation for the variety. Thus *<GY, <PTH, and <RGH* represent Guoyu, Putonghua, and Rugaohua, respectively. The label is placed at the point where the speaker starts to use the variety in question.

More complex labels can also be created to identify the particular feature(s) involved in a partial code-switch. This is illustrated in Figure 9.1(b). The utterance is a reading in Rugaohua of a text originally composed in Guoyu. Most of the lexical items have everyday cognates in Rugaohua, but *háizimen* is a word 'borrowed' from the standard variety of Mandarin (which for the Rugaohua speaker would be Putonghua rather than Guoyu). Since she was reading a prepared text, the speaker produced the Putonghua word *háizimen* using Rugaohua tones, rather than substituting the native Rugaohua word for 'children'. On the code tier in the figure, the sentence-initial <RGH marks the variety of Mandarin used to produce the utterance. The <PTH-word label then marks the beginning of the standard Mandarin lexical item, and <RGH signals the point where the speaker resumes a pure Rugaohua code.

9.4. CONCLUSION

In this chapter, we have described some of the salient prosodic phenomena of Mandarin, and proposed the development of a Pan-Mandarin M_ToBI system (see Appendix for a summary of the proposed annotations). Much research will be required before we have a complete system that can cover all three national standards and at least a few of the major regional varieties such as those spoken in Rugao, Xi'an, Chengdu, Nanjing, and so forth. It will probably be several years before we can even say with confidence that the broad outline of the overall structure of tiers is adequate to cover all varieties. Mandarin is spoken over a vast geographical area. Different regional dialects abut other language varieties that range from Shanghai (a Wu variety of Chinese) in the east to Tibetan (a Tibeto-Burman language) and Mongolian (an Altaic language) in the extreme west and north of the Mandarin-speaking region of the PRC. The range of prosodic variability is probably much larger than that for Cantonese (see Wong *et al.*, this volume Ch. 10).

However, even at this preliminary stage, it is clear that developing a Pan-Mandarin M_ToBI system gives us a manageable way to frame the questions for future research. For example, the separation of degrees of perceived juncture (on the break indices tier) from phrasal pitch range effects (on the tones tier) allows us to explore the first of these in inter-transcriber calibration tests without committing prematurely to categorical definitions of prosodic units in a strictly layered hierarchy. The separation gives us a theory-neutral way to explore the potentially quite variable relationship across different styles and different varieties between prosodic grouping or stress, on the one hand, and tonal reduction versus tone sandhi, on the other.

It is also clear that the ToBI framework will need to be extended in at least two ways to apply to Mandarin. First, unless we limit the scope of M_ToBI to Guoyu (or perhaps to just the three national standards), we will need a device such as the code tier to accommodate to prosodic differences across varieties and to describe phenomena such as tonal code switching. Second, because Mandarin is a 'stress' language, and because even the standard varieties differ markedly on the distribution of unstressed syllables relative to words and phrases, stress needs to be tagged explicitly. For Mandarin then, the ToBI framework probably should become the ToBISL framework ('Tones, Break Indices, and Stress Levels').

APPENDIX: SUMMARY OF M_ToBI LABELS

Label	Tier	Description
S3	Stress	Syllable with fully-realized lexical tone.
S2	Stress	Syllable with substantial tone reduction.
S1	Stress	Syllable that has lost its lexical tonal specification.
S0	Stress	Syllable with lexical neutral tone.
35	Sandhi	Tone 3 realized as a sandhi tone (rising tone).
214	Sandhi	Tone 3 realized as a low dipping tone.
H%	Tones	High boundary tone (at the end of an utterance).
L%	Tones	Low boundary tone (at the end of an utterance).
%reset	Tones	Beginning of a new pitch downtrend or pitch reset.

%q-raise	Tones	Beginning of a raised pitch range.
%e-prom	Tones	Beginning of local expansion of pitch range due to emphatic prominence.
%compressed	Tones	Beginning of reduction of pitch range of syllables following the expansion of pitch range under %e-prom.
B0	Break Indices	Reduced syllable boundary, i.e. contraction: requires S_0 or S_1 on left or right.
B1	Break Indices	Normal syllable boundary: the 'default' case within a polysyllabic word.
B2	Break Indices	Minor phrase boundary: must be followed by at least S_2.
B3	Break Indices	Major phrase boundary.
B4	Break Indices	Breath group boundary: reset of pitch between sentences or phrases (H%, L%, %reset, and %q-raise on the tones tier).
B5	Break Indices	Prosodic group boundary: a breath group boundary accompanied by a prolonged pause.
<	Code	Beginning of the variety of Mandarin used: followed by the abbreviation for the variety, e.g., < GY for Guoyu; < PTH for Putonghua.

REFERENCES

Arvaniti, A., and Balthazani, M. (this volume Ch. 4), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora'.

Beattie, D. A. (1985), 'Mandarin Tone Sandhi and the Interaction of Phonology and Syntax', thesis, University of Toronto, Toronto.

Bruce, G. (this volume Ch. 15), 'Intonational Prominence in Varieties of Swedish Revisited'.

Cassidy, S., and Harrington, J. (2001), 'Multi-level Annotation in the Emu Speech Database Management System', *Speech Communication*, 33/1–2: 61–77.

Chan, M. K.-M. (1991), 'Contour-tone Spreading and Tone Sandhi in Danyang Chinese', *Phonology* 8: 237–59.

——, and Ren, H.-M. (1989), 'Wuxi Tone Sandhi: from Last to First Syllable Hominance', *Acta Linguistica Hafniensia*, 21/2: 35–64.

——, and Tai, J. H-Y. (1989), 'A Critical Review of Norman's *Chinese*', *Journal of the Chinese Language Teachers Association*, 24/1: 43–61.

Chao, Y.-R. (1930), 'A System of Tone Letters', *Le Maitre Phonetique*, 30: 24–7, repr. in regular English orthography in *Fangyan* (1980) 2: 81–3.

—— (1933), 'Tone and Intonation in Chinese', *Bulletin of the Institute of History and Philology*, 4: 121–34.

—— (1948), *Mandarin Primer* (Cambridge, MA: Harvard University Press).

—— (1968), *A Grammar of Spoken Chinese* (Berkeley: University of California Press).

Chen, M.-Y. (2000), *Tone Sandhi: Patterns Across Chinese Dialects* (Cambridge, UK: Cambridge University Press).

Cheng, C.-C. (1973), *A Synchronic Phonology of Mandarin Chinese* (The Hague: Mouton).

Cheng, R.-L. (1985), 'A Comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin', *Language*, 61/2: 352–77.

Chou, F.-C., Tseng, C.-Y., and Lee, L.-S. (1996), 'Automatic Generation of Prosodic Structure for High Quality Mandarin Speech Synthesis', *Proceedings of the Fourth International Conference on Spoken Language Processing* (Philadelphia, PA), 1624–7.

——, ——, —— (1998), 'Automatic Segmental and Prosodic Labelling of Mandarin Speech Database', *Proceedings of the Fifth International Conference on Spoken Language Processing* (Sydney, Australia).

Duanmu, S. (1990), 'A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages', Ph.D. dissertation (Massachusetts Institute of Technology) (distributed by MIT Working Papers in Linguistics).

—— (1997), 'Recursive Constraint Evaluation in Optimality Theory: Evidence from Cyclic Compounds in Shanghai', *Natural Language and Linguistics Theory*, 15: 465–507.

Feifel, K.-E. (1994), *Language Attitudes in Taiwan: A Social Evaluation of Language in Social Change* (Taipei: Crane Publishing).

Fon, J. (2000), 'Production and Perception of Two Dipping Tones (T2 and T3) in Taiwan Mandarin', ms, Department of Linguistics, Ohio State University.

——, and Chiang, W.-Y. (1999), 'What does Chao have to say about tones?—a case study of Taiwan Mandarin', *Journal of Chinese Linguistics*, 27/1: 15–37.

Grice, M., D'Imperio, M., Savino, M., and Avesani, C. (this volume Ch. 13), 'Strategies for Intonation Labelling Across Varieties of Italian'.

Grimes, B. F. (1996), (ed.), *Ethnologue: Languages of the World* (13th edn., Dallas, Texas: Summer Institute of Linguistics) (the statistics on Mandarin speakers are from the online version < http://www.sil.org/ethnologue/ >).

Hartman, L. M. (1944), 'The Segmental Phonemes of the Peiping Dialect', *Language*, 20: 28–42.

Hsiao, Y.-C. E. (1991), *Syntax, Rhythm and Tone: A Triangular Relationship* (Taipei: Crane Publishing).

Hu, M.-Y. (1987), 'Putonghua he Beijinghua' [Putonghua and Beijing Dialect], *Beijinghua Chu Tan* [A Preliminary Study of the Beijing Dialect] (Beijing: Commercial Press) repr. in Hu, M.-Y., *Yuyanxue Lunwenji* [Selected Writings in Linguistics] (Beijing: Zhongguo Renmin Daxue Chubanshe, 1991), 167–87.

Huang, T. (1999), 'A First Look at Rugao Chinese', paper presented at the Colloquiumfest, Department of Linguistics, Ohio State University.

Hung, T. T.-N. (1989), 'Syntactic and Semantic Aspects of Chinese Tone Sandhi', Ph.D. dissertation (University of California, San Diego) (reproduced by the Indiana University Linguistics Club, Bloomington, IN).

Jin, S.-D. (1985), 'Shanghai Morphotonemics: A Preliminary Study of Tone Sandhi Behavior Across Word Boundaries', thesis, University of Pittsburgh (reproduced by Indiana University Linguistics Club, Bloomington, IN).

—— (1996), 'An Acoustic Study of Sentence Stress in Mandarin Chinese', Ph.D. dissertation (Ohio State University).

Kratochvil, P. (1968), *The Chinese Language Today: Features of an Emerging Standard* (London: Hutchinson University Library).

—— (1987), 'The Case of the Third Tone', in the Chinese Language Society of Hong Kong (eds.), *Wang Li Memorial Volumes, English Volume* (Hong Kong: Joint Publishing Co., 253–77).

—— (1998), 'Intonation in Beijing Chinese', in D. Hirst and A. Di Cristo (eds.) *Intonation Systems: A Survey of Twenty Languages* (New York: Cambridge University Press), 417–31.

Lee, O. J. (2000), 'The Pragmatics and Intonation of Ma-Particle Questions in Mandarin', thesis (Ohio State University).

Li, A.-J., Zu, Y.-Q., and Li, Z.-Q. (1999), 'A National Database Design and Prosodic Labelling for Speech Synthesis', paper presented at the 1999 Oriental COCOSDA Workshop on East Asian Language Resources and Evaluation, Taipei, Taiwan, 13 May.

Li, R. (1985), 'Guanhua Fangyan de Fenqu' [The Grouping of the Mandarin Dialects], *Fangyan*, 1: 2–5.

—— (1989*a*), 'Zhongguode Yuyan he Fangyan' [Languages and Dialects in China], *Fangyan*, 3: 161–7.

—— (1989*b*), 'Hanyu Fangyande Fenqu' [The Classification of the Chinese Dialects], *Fangyan*, 4: 241–59.

Liao, R.-R. (1994), 'Pitch Contour Formation in Mandarin Chinese: A Study of Tone and Intonation', Ph.D. dissertation (Ohio State University).

Liberman, M., and Pierrehumbert, J. (1984), 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aronoff, and R. Oerhle (eds.), *Language and Sound Structure* (Cambridge, MA: MIT Press), 157–233.

Peng, S.-H. (1996), 'Phonetic Implementation and Perception of Place Coarticulation and Tone Sandhi', Ph.D. dissertation (Ohio State University).

Qian, Z.-Y. (1982), *Yantai Fangyan Baogao* [Report on the Yantai Dialect] (Ji'nan: Qilu Shushe).

Qingdaoshi Shizhi Bangongshi (1997), *Qingdao Shizhi: Fangyan Zhi.* [Qingdao City Records: Dialect Records] (Beijing: Xinhua Chubanshe).

Selkirk, E., and Shen, T. (1990), 'Prosodic Domains in Shanghai Chinese', in S. Inkelas and D. Zec (eds.), *The Phonology–Syntax Connection* (Chicago: University of Chicago Press), 313–37.

Shen, X.-N. (1989), 'Interplay of the Four Citation Tones and Intonation in Mandarin Chinese', *Journal of Chinese Linguistics*, 17/1: 61–74.

—— (1990), *The Prosody of Mandarin Chinese* (Berkeley: University of California Press).

Shih, C.-L. (1986), 'The Prosodic Domain of Tone Sandhi in Chinese', Ph.D. dissertation (University of California, San Diego).

—— (1988), 'Tone and Intonation in Mandarin', *Working Papers of the Cornell Phonetics Laboratory*, 3: 83–109.

—— (1997), 'Mandarin Third Tone Sandhi and Prosodic Structure', in J.-L. Wang and N. Smith (eds.), *Studies in Chinese Phonology* (Berlin and New York: Mouton de Gruyter), 81–123.

Stimson, H. M. (1966), 'A Tabu Word in the Peking Dialect', *Language*, 42/2: 285–94.

Tai, J. H.-Y. (1976), *Lexical Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, DC: Office of Research, United States Information Agency).

—— (1977), *Syntactic and Stylistic Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, DC: Office of Research, United States Information Agency).

—— (1978), *Phonological Changes in Modern Standard Chinese in the People's Republic of China Since 1949* (Washington, DC: Office of Research, United States Information Agency).

Ting, P.-H. (1966), 'Rugao Fangyan de Yinyun' [Phonology of the Rugao Dialect], *Bulletin of the Institute of History and Philology*, 36: 573–633.

Tseng, C.-Y, and Chou, F.-C. (1999*a*), 'Machine Readable Phonetic Transcription System for Chinese Dialects Spoken in Taiwan', *Journal of Acoustical Society of Japan (E)* 20/3: 215–23.

——, —— (1999*b*), 'A Prosodic Labeling System for Mandarin Speech Database', *Proceedings of the XIVth International Congress of the Phonetic Sciences*, Vol. 3, 2379–82.

Venditti, J. (this volume Ch. 7), 'The J_ToBI Model of Japanese Intonation'.

Wang, W. S.-Y., and Li, K.-P. (1967), 'Tone 3 in Pekingese', *Journal of Speech and Hearing Research*, 10: 629–36.

Wong, P. W.-Y., Chan, M. K.-M., and Beckman, M. E. (this volume Ch. 10), 'An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese'.

Wright, M. (1983), 'A Metrical Approach to Tone Sandhi in Chinese Dialects', Ph.D. dissertation (University of Massachusetts).

Xu, Y. (1999), 'Effects of Tone and Focus on the Formation and Alignment of Fo Contours', *Journal of Phonetics*, 27/1: 55–105.

Yin, Z.-Y. (1982), 'Guanyu Putonghua Shuangyin Changyongci Qingzhongyinde Chubu Kaocha' [A Preliminary Study of Accents and Atonics in Disyllabic Words in Common Use], *Zhongguo Yuwen*, 3: 168–73.

Yip, M. (1980), 'The Tonal Phonology of Chinese', Ph.D. dissertation (Massachusetts Institute of Technology).

Yuan, J.-H. (1960), *Hanyu Fangyan Gaiyao* [Outline of Chinese Dialects] (Beijing: Wenzi Gaige Chubanshe) (second edition published in 1989).

Zee, E. (1980), 'A Spectrographic Investigation of Mandarin Tone Sandhi', UCLA Working Papers 49: 98–116.

Zhang, Z.-S. (1988), 'Tone and Tone Sandhi in Chinese', Ph.D. dissertation (Ohio State University).

10

An Autosegmental-Metrical Analysis and Prosodic Annotation Conventions for Cantonese

Wai Yi P. Wong, Marjorie K. M. Chan, and Mary E. Beckman

10.1. INTRODUCTION

The Hong Kong Cantonese variety of Chinese (hereafter 'Cantonese') poses an interesting challenge for prosodic typology and transcription for three closely inter-related reasons. First, compared to the Mandarin varieties of Chinese, Cantonese has far fewer polysyllabic word forms. The majority of the syllables are potentially free-standing morphemes, and there is no contrast between 'stressed' syllables and reduced ('neutral-tone') syllables. Second, there is an extremely dense syntagmatic specification of tone. Every syllable in an utterance has a lexical tone, even if it is a grammatical morpheme or pragmatic particle, and there is a rich inventory of non-segmental pragmatic morphemes ('boundary tones') that can be added after the final lexical tone to mark the ends of intonational phrases. Finally, while there are occasional consonant lenitions, vowel deletions and attendant resyllabifications that create 'fused forms' of familiar phrases in running speech, there seem to be no other reliable categorical markings of intermediate levels of prosodic grouping between the syllable and the intonational phrase. Thus, it is difficult to define a low-level unit comparable to the 'prosodic word' of Greek (Arvaniti and Baltazani this volume Ch. 4), the 'accentual phrase' of Korean (Jun this volume Ch. 8), or the 'tone sandhi group' in the Wu varieties of Chinese (Jin 1985; Selkirk and Shen 1990). The C_ToBI (*C*antonese *T*ones and *B*reak *I*ndices) conventions are

We thank Steve Cassidy and Jonathan Harrington of the Speech and Hearing Research Centre at Macquarie University for advice in the implementation of the EMU speech database labelling platform at our site. We also thank Sun-Ah Jun and an anonymous reviewer for providing us with thoughtful suggestions and comments.

designed to annotate and explore these tone and juncture phenomena in spoken language corpora. They are developed within the Autosegmental-Metrical approach of the ToBI framework, which represents the string of tones on its own autosegmental tier, independent of the different structural positions that license tones at different levels of the metrical hierarchy.

While the development of the C_ToBI annotation conventions is based primarily on modern Hong Kong Cantonese, the proposal is for common levels of transcription for C_ToBI users at different sites for transcribing other varieties of Cantonese, such as Guangzhou (Canton) Cantonese, Zhongshan Cantonese, and so forth. The three properties that we outlined above as characteristic of Hong Kong Cantonese seem to hold in broad terms also for these other varieties. Variation that we have noticed in our own casual observations seems to involve only some paradigmatic differences in lexical tone and boundary tone inventory (and not in the dense syntagmatic specification of tone) as well as differing propensities for fused forms. (Fused forms seem to occur rather less frequently and only in more casual styles in less urban varieties.) Also, at this preliminary stage, development has been based primarily on constructed examples, although work is also in progress on an elicited spontaneous speech corpus. Again, however, it seems unlikely that further work with more different styles will change the broad outline of facts outlined above. In designing C_ToBI, therefore, we have tried to capture the generally shared structure, while leaving scope for change in detail as we compare transcriptions across different styles, and especially across different transcription sites. Cross-site comparison of transcriptions is desirable in order to arrive at an agreed-upon set of levels of transcription congenial to the community at large. Only in this way can we hope to be able to capture any sociolinguistically significant variation in prosody across the different varieties of Cantonese and across the different styles that a native speaker of Cantonese controls in everyday life.

The challenge that Hong Kong Cantonese poses for prosodic annotation then comes from several facts that seem to be true of Cantonese in general. Moreover, these facts about Cantonese prosody make it different not just from languages such as American English and Tokyo Japanese, but also from other major varieties of Chinese, including Mandarin (as described in Peng *et al.* this volume Ch. 9). We will give an account of these facts in Section 10.2. To capture these essential aspects of the language, we take advantage of the facilities that the ToBI framework has provided—namely, that it suggests only a basic differentiation between tonal specification and the prosodic groupings with which tones can be associated. Beyond this distinction between melody (*To*nes) and junctures (*B*reak *I*ndices), it allows for free proliferation of tiers and user-defined labels specific to the needs of the language in question and the interests of individual sites. To illustrate, we will lay out details of the C_ToBI annotation conventions in Section 10.3. The conventions were developed out of what we know already about the prosody of Cantonese and our initial hypotheses that require further testing. In Section 10.4, we will close by listing some questions of interest that we plan to explore using C_ToBI.

10.2. FACTS ABOUT CANTONESE PROSODY

10.2.1. *What do we mean by 'Cantonese'?*

We use the term 'Cantonese' in three senses. The first and oldest sense is in reference to the variety of Chinese spoken in and around the city of Canton (Guangzhou), the provincial capital of Guangdong Province in southern China. Because many Canton City inhabitants have migrated to Hong Kong in the past 150 years or so (Yue-Hashimoto 1972: 70), the variety of Cantonese spoken in Hong Kong is more similar to that of Canton than are varieties of the language spoken in towns closer to Canton City itself. Hence, socially prestigious 'Standard Cantonese' can refer to the Cantonese spoken in Hong Kong as well as in Canton, even though over time a number of differences have emerged in the speech of these two localities (Bauer and Benedict 1997). It is in this second usage of the term 'Cantonese' that we focus here on modern Hong Kong Cantonese. Finally, the term 'Cantonese' has further been extended to refer to the entire group of similar dialects that, together, form one of the major sub-classifications within Chinese, parallel to Mandarin, Wu, Min, and so forth. (One also encounters the term 'Yue' for that usage—see, for example, Yue-Hashimoto 1972; Yuan 1983; Grimes 1996.) Because of its close affiliation with the original regional standard of Canton City, the Standard Cantonese of Hong Kong can also serve as representative of that major sub-group within the Chinese language. However, it is important to recognize that many other varieties of Cantonese in this third sense are also found in Hong Kong and vicinity, and the present-day system cannot be understood without taking into account the rich sociolinguistics of contact among these varieties.

According to a 2001 report from the Hong Kong Census and Statistics Department, of the 6,708,389 inhabitants in Hong Kong, 89.2 per cent speak Cantonese as their mother tongue, which is also the language of everyday use. Cantonese is also found in numerous overseas speech communities around the world, including Singapore, Malaysia, Vietnam, Indonesia, Thailand, Philippines, New Zealand, United Kingdom, United States, and Canada. In fact, the Summer Institute of Linguistics' *Ethnologue: Languages of the World* (Grimes 1996) gives an estimate of some 66 million speakers of Cantonese ('Yue') in the world today, and ranks it as sixteenth in its list of top 100 languages by population. Ideally, we would like C_ToBI to be able to account for the prosodic structure across this world-wide community of Cantonese speakers. At the least, it should describe the situation in Guangzhou, Hong Kong, and adjacent Cantonese-speaking areas.

10.2.2. *The syllable*

One important characteristic of Cantonese is the highly salient psychological reality of the syllable. The syllable is easy to define at the phonological level. The dense syntagmatic distribution of lexical tones marks the number of syllables in any stretch of speech more clearly than in any other variety of Chinese. Also, outside of fused forms, syllable boundaries are clearly identifiable from asymmetries in the distribution of onset and coda segments. That is, there is nothing like the large-scale ambiguities of segmentation within the stress feet of an English utterance. The morphology of the language is also highly conducive to thinking phonologically in terms of syllables, since more than any other major variety of Chinese, Cantonese is probably closest to the original Chinese morphological structure of a strong one-to-one correspondence between words and syllables.

The Cantonese syllable has an optional onset consonant and a rhyme, which (in addition to the obligatory tone specification) consists of either a simple vowel, a simple vowel followed by an optional coda consonant, a vowel-glide diphthong, or syllabic nasal. While there is a fairly rich inventory of consonants in onset position—including labial, dental, and velar nasals, and labial, dental, velar, and labiovelar stops in both a voiceless aspirated and a voiceless unaspirated series—consonants in coda position are restricted to the three nasals and labial, dental, and velar voiceless (unreleased) stops. The vowel inventory includes both front and back rounded vowels. The full inventory of consonants and vowels in Hong Kong Cantonese is given in Appendix I. At the level of the syllable, there are some co-occurrence restrictions. Most notably, labials and labiovelars almost never occur as onsets in syllables that have either labial codas or rounded vowels (Yue-Hashimoto 1972: 110ff, 137ff; see also Newman 1987; Yip 1988; Bauer and Benedict 1997), with other places showing less robust Obligatory Contour

Principle (OCP) effects (cf. Wong 2002*a*). The rare exceptions to the labial OCP include onomatopoeia and loanwords.

10.2.3. *Tone specification*

As noted in Section 10.1 above, different varieties of Cantonese differ in their lexical tone inventories, and 'tone change' is one of the more salient variables in the sociolinguistic mix (see Section 10.2.5). Hong Kong Cantonese today has the six phonemic tones illustrated in Figure 10.1 on the syllable /wai/. These tones are traditionally described using Yuan-ren Chao's system of tone numbers, with '1' for lowest pitch and '5' for the highest pitch in the local pitch range. There are three level tones (/55/, /33/, /22/), two rising tones (/35/, /23/), and one falling tone (/21/).[1] Some Cantonese speakers also have a high falling tone (/53/), which for most Hong Kong speakers has merged with the high level tone. In checked syllables (i.e. syllables closed by a stop) the three level tones are traditionally transcribed with single digits (i.e. /5/, /3/, /2/), iconically reflecting the shorter duration of the vowel. The C_ToBI labelling conventions adopt this traditional transcription of lexical tone, except that the iconic encoding of the durational difference between checked and non-checked syllables is made uniform by the addition of an 'extra' digit to the rising and falling tones in syllables with all-sonorant rhymes, as shown in

[1] The choice of /35/ versus /23/ for the two rising tones follows Chao (1947). Note, however, that the literature on Cantonese is not unanimous in assigning these specific descriptors. Bauer and Benedict (1997: 144), for example, argue for /25/ as the transcription of the mid-rising tone in modern Hong Kong Cantonese, on the grounds that the tone now starts at a mid-low level rather than at the middle of the speaker's pitch range. Bauer (personal communication) suggests further that Chao's transcription is of an older variety that distinguished an underlying /35/ from a morphologically conditioned 'changed tone' (bianyin). That is, the lowered onset may be the residue of a neutralizing sound change. Our own general observation is that speakers differ in their production of the two rising tones, especially in uttering the citation tones in sequence using the same syllable. For some speakers, the mid-rising tone is produced with overall higher pitch than the low-rising tone. For other speakers, however, it seems that the two rising tones have a similar starting point—mid-low—and where they differ is in slope and final tonal target; that is, the mid-rising tone rises much more sharply and ends much higher in pitch than the low-rising tone. The differences may be due in part to sociolinguistic factors, a possibility that requires further research. The C_ToBI system is intended to promote such research, by providing a means to tag phonological contrasts in spoken language corpora which are large and varied enough to provide good statistical control of sociolinguistic and discourse factors that can influence pitch range and target. That is, terms such as '/35/' should not be interpreted as phonetic transcriptions. They cannot substitute for the fundamental frequency trace or other comparably fine-grained phonetic representations. Rather /35/ versus /23/ should be understood merely as mnemonic tags into the actual speech data—a tool for keeping track of the phonological categories that are relevant when extracting finer-grained phonetic detail from quantitative models of all of the factors that affect fundamental frequency.

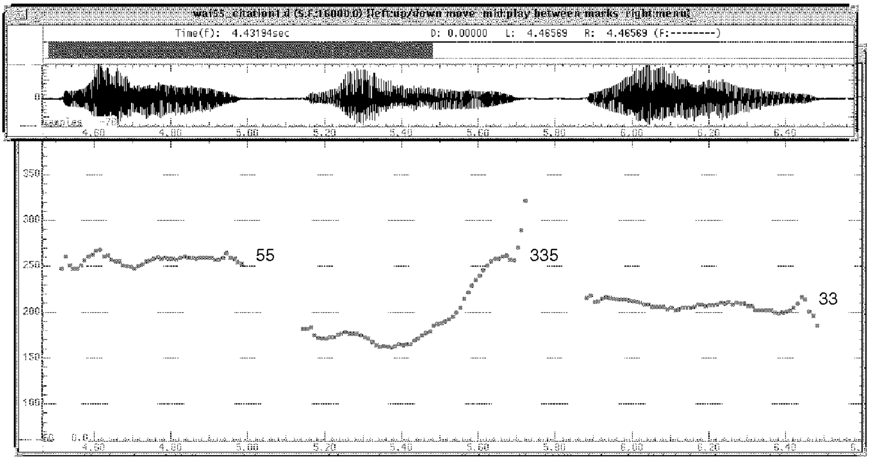


FIGURE 10.1(a) A display of three of the six phonemic tones in modern Hong Kong Cantonese /55, 335, 33/, as illustrated by the syllable /wai/ (in Jyutping romanization, see Appendix I for the scheme) with Fo traces. Annotated at the end of each tone are Chao-type tone numbers as modified in the C_ToBI labelling conventions, with '1' to '5' in ascending pitch height (e.g., /55/ is high-level; /221/ is low falling, etc.) (cf. Table 10.1).

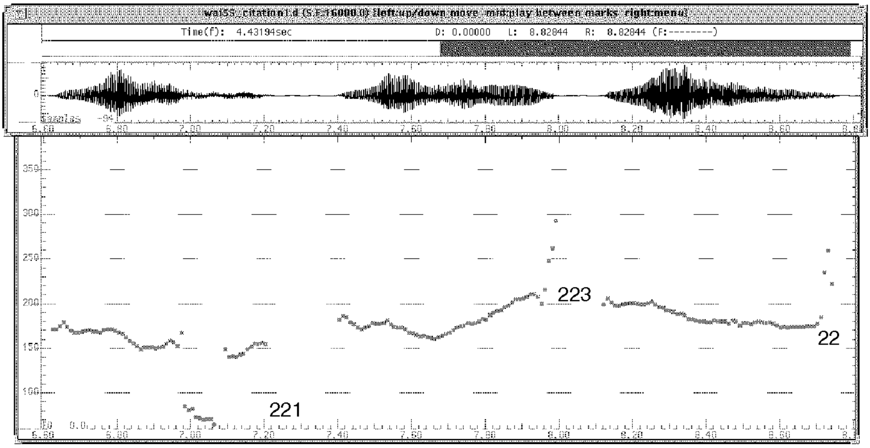


FIGURE 10.1(b) A display of three of the six phonemic tones in modern Hong Kong Cantonese /221, 223, 22/, as illustrated by the syllable /wai/ with Fo traces. Annotated at the end of each tone are Chao-type tone numbers as modified in the C_ToBI labelling conventions, with '1' to '5' in ascending pitch height (e.g., /55/ is high-level; /221/ is low falling, etc.) (cf. Table 10.1).

Figure 10.1. (More detail on the labelling conventions for lexical tone is given in Section 10.3.1(iii).)

Cantonese also has boundary tones. These are 'extra' tones that can be added after the final lexical tone of an intonational phrase to produce various pragmatic effects. The C_ToBI conventions distinguish boundary tones from the lexical tones by transcribing them with labels that (following Pierrehumbert 1980) have become standard in autosegmental-metrical descriptions of other languages. Figures 10.2 to 10.4 illustrate three of the boundary tones of Hong Kong Cantonese, all added at the end of an utterance after a /33/ lexical tone on the utterance-final syllable. Observe the substantial lengthening of the final word, particularly for the complex boundary tone in Figure 10.4, where one could almost say that an extra lengthening is necessary to make room for these two 'extra' tones at the phrase boundary. Figure 10.3 also illustrates an utterance-medial phrase boundary where there is substantial final lengthening, but no 'extra' tone after the mid-level lexical tone on the particle /aa33/.

The boundary tones that we have observed so far in our recordings of Hong Kong Cantonese are inventoried further in section 10.3.1(iii). As with the lexical tones, we expect to discover variation and change in the inventory

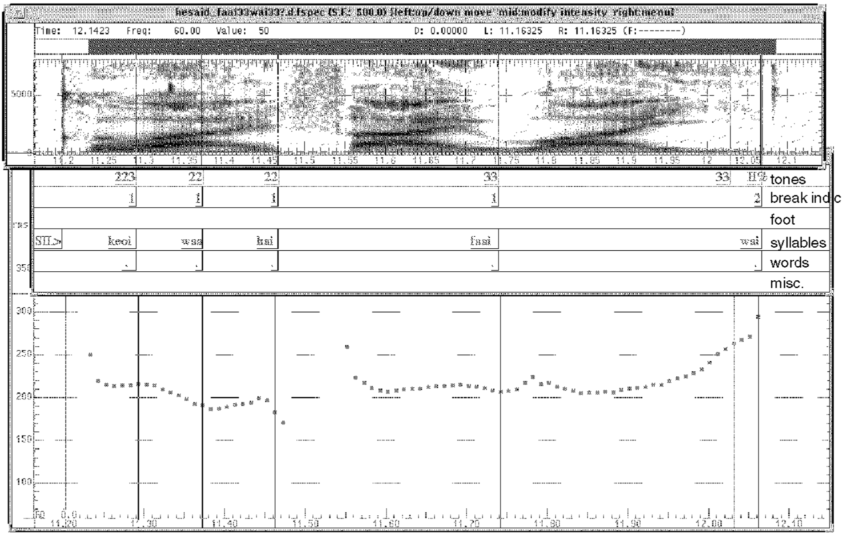


FIGURE 10.2 Fo contour of the utterance '*Keoi223 waa22 hai22 faai33 wai33? (S/he said it was (the word) "pleasant"?)*', with the pragmatic boundary tone H% attached to the final tone, indicating seeking confirmation, with probably a connotation of surprise (figure transcribed in C_ToBI; function of each tier of transcription in the C_ToBI system is noted on the right margin of the tiers).

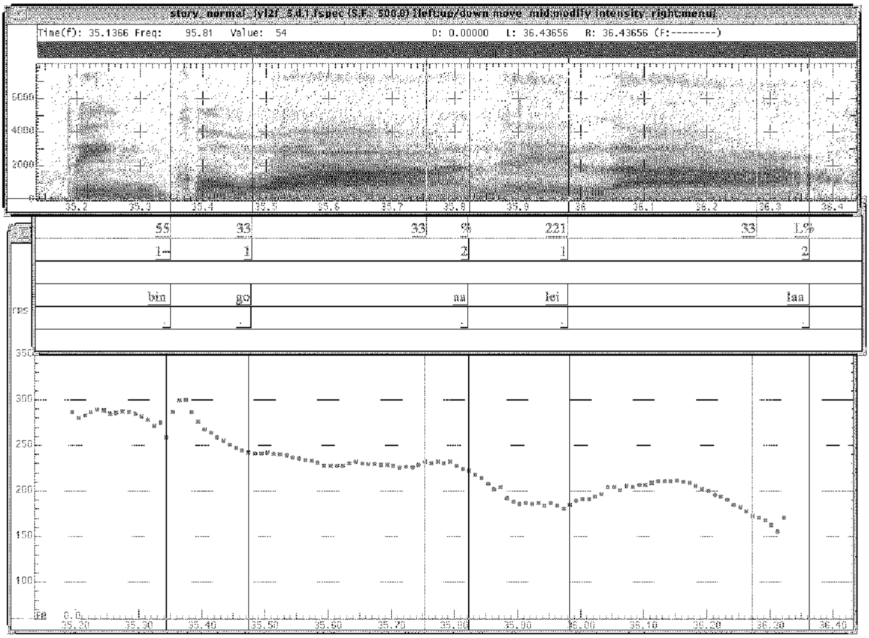


FIGURE 10.3 F0 contour of two prosodic phrases '*Bin55 go33 aa33[final particle]? Lei221 laa33[final particle]. (Who is it? Coming.)*' marked by two phrase boundaries. Notice that the final tone of the second utterance is attached with the pragmatic boundary tone L%. Demonstrated here is a declarative use of L% (figure transcribed in C_ToBI).

of boundary tones across varieties of Cantonese spoken in different regions and by different social groups. However, we do not expect varieties to differ from Hong Kong Cantonese in not having boundary tones. These tones and the other effects with which they are associated (such as the 'phrase-final lengthening' just described) thus define a level of prosodic grouping above the word. This level is aptly termed the 'intonational phrase' in keeping with the comparable level of prosodic hierarchy in German, Japanese, Greek, etc. (see Chapters 3, 7 and 4 by Grice, Baumann, and Benzmüller; Venditti; and Arvaniti and Baltazani respectively).

As Figures 10.2 to 10.4 well illustrate, Hong Kong Cantonese has a very dense syntagmatic specification of tone, with each syllable in a well-formed phrase bearing a lexical tone and the phrase-final syllable also often bearing a boundary tone that is a separate pragmatic morpheme specified for the phrase as a whole. Thus, each syllable is specified for one of the six tones in the language, and pragmatic particles are no exception. That is, by contrast to Taiwanese, Mandarin, and probably most other varieties of Chinese, Cantonese particles

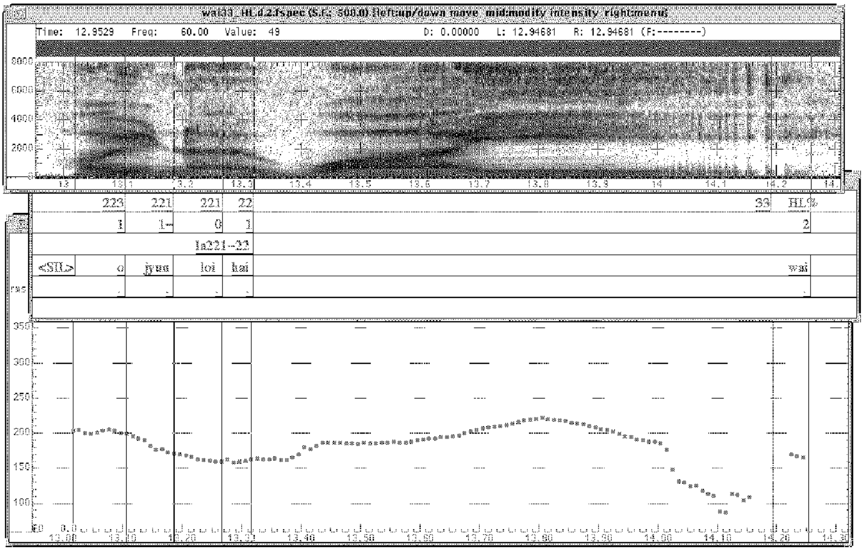


FIGURE 10.4 F0 contour of the utterance '*O223 jyun221 loi221 hai22 wai33. (Oh, I see, so it was (the word) "fear".)*', with the pragmatic boundary tone HL% attached to the final tone. This use of HL% is a typical way to express connotation of 'discovery' in Hong Kong Cantonese (figure transcribed in C_ToBI).

bear lexical tone. There are no unstressed (neutral tone) syllables, not even for these 'little' function words. Cantonese has a particularly rich inventory of final particles. Depending on the treatment of closely related variant forms and of more or less conventionalized particle sequences, the count can be anywhere from about thirty (Kwok 1984) to as many as 206 (Yau 1980). This is in marked contrast to the seven particles commonly listed for Standard Mandarin (e.g. Matthews and Yip 1984: 238). Since these final pragmatic particles are by definition final, they interact with tonal pragmatic morphemes ('boundary tones') to yield very complex pragmatic effects (cf. Kwok 1984; Chan *et al.* 1998; Fung 2000). Precisely how these final pragmatic particles and tonal pragmatic morphemes interact leaves ample room for further research. The development of the C_ToBI annotation conventions should contribute a tool to help us understand the precise nature of the interaction.

10.2.4. *Fusion forms*

There are segmental effects in Hong Kong Cantonese that fuse two syllables together into a polysyllabic 'word' in fast speech (Li 1986; Wong 1996;

Wong 2002*b*) (see example (1b)). Thus, the facts about syllable structure outlined above in Section 10.2.2 must be qualified to some extent, at least for the Hong Kong variety of Cantonese. This phenomenon of 'fusion' has not been systematically studied. Indeed, there is little prior work on any aspect of Cantonese connected speech. Several examples of fused forms are listed in (1). (See also examples in Figures 10.4 to 10.7 in Section 10.3.1(v).)

We define syllable fusion as follows: in a sequence of two syllables, there is substantial weakening or effective deletion of the oral gesture(s) of the segment(s) contiguous to the syllable boundary. More extreme fusion can simplify contour tones and 'merge' the qualities of vowels that would be separated by an onset or coda consonant at more 'normal' degrees of disjuncture between words. However, even at these more extreme junctures, fusion does not usually override the lexical tones of the syllables (Yue-Hashimoto 1972; Cheung 1986; Li 1986; Wong 1996; Wong 2002*b*), and it is only in the most extreme cases (e.g. *mat5 je23* 'what' → *me55* 'what') where we see true 'tone loss'.[2] The fusion process, therefore, seems to be a gradient effect. Although it is possible that the most extreme cases of fusion may make the syllable count less determinate, fusion does not seem to be a categorical change from two syllables to one in all cases, contra our own earlier description in Wong, P. W.-Y. (1996).

(1) Examples of syllable fusion forms.
(*a*) know NEG
sik5.m21 → siʔ5.m21 → si5.m21
(*b*) in fact
kei21.sat2 → ke21.at2 → ket21-2
(*c*) know
zi55.dou33 → zi55.ou33 → ziu55-33

Li (1986) describes fusion as the postlexical counterpart of historical contractions resulting in 'changed tones' (see next section). A complete account of fusion probably will need to place it in the context of the sociolinguistics of contact among these dialects. Our current preliminary account is in terms of the foot (cf. also Yip 1980; Duanmu 1990). That is, we ascribe the 'monosyllabic' flavour of Cantonese to the preponderance of monosyllabic feet. Fused forms then are exceptions to the regular one-to-one

[2] Note that in these cases, the fusion typically results in lexicalization of the contracted form. That is, most native speakers probably would identify *me55* 'what' as a separate lexical item.

correspondence between foot and syllable, and may presage a change toward a system with a more well-defined intermediate level of prosodic grouping between the syllable and intonational phrase.

10.2.5. *Segmental and tonal alternations*

In addition to the gradient variation that results from fusion, there are several categorical segmental and tonal alternations in the Hong Kong variety of Cantonese that are of interest in the development of C_ToBI, for two reasons. First, some of these alternations must have arisen historically through the lexicalization of connected speech processes and discourse-level phenomena described in the sections above, and thus give a hint of the kinds of prosodic change in progress to look for in modern Cantonese. Second, many of these alternations are socially meaningful consequences of the extreme linguistic diversity that has characterized Hong Kong as long as it has existed. (The territory was ceded to England and opened as a port city in 1842.)

Hong Kong's history of rural–urban contact, with an accelerated pace of urban immigration in the last half century, has made it a mixing pot for Cantonese dialects, much like Shanghai in relationship to the Wu varieties much earlier. For example, in addition to Hong Kong Cantonese, other varieties encountered in Hong Kong include a number of dialects (e.g., Dongguan, Panyu, Shunde, Taishan, Kaiping) in which velars are merging with dentals in coda position (cf. Zhan and Cheung 1987; Zee 1999*b*). The resulting alternation between /t, n/ and /k, ng/ obscures further the less robust OCP effects with lingual consonants. The role of English as an administrative language further increased the linguistic diversity (see, for example, the stratum of loanwords from English as exceptions to the labial OCP constraint). Nor is English the only non-Cantonese source of potential influence on Hong Kong Cantonese. There have always been speakers of other varieties of Chinese, such as Chiuchao (Chaozhou), Fukien (Fujian), Hakka (Kejia), Shanghainese, and Putonghua Mandarin (Hong Kong Census and Statistics Department 1996: 22). With the change of government in 1997, there has been a further influx of such non-Cantonese Chinese speakers. And the trend is likely to increase with the implementation of an immigration regulation in 2000 that would permit mainland Chinese to obtain approval of residency in Hong Kong at the rate of around 156 heads per day.

Examples of some of the segmental alternations found in Hong Kong Cantonese are shown in (2). (See Yue-Hashimoto 1972; Bauer 1979; Yeung 1980; Rao *et al.* 1981; Bourgerie 1990; Matthews and Yip 1994 for more examples.) These alternations occur on consonants in both onset and coda positions, as well as on the nuclear vowel. They may be conditioned lexically or occur across the board. These alternations are often said to be related to more casual versus more formal styles of speech. However, the relationship is not clear-cut. Much work needs to be done before we can adequately characterize the interaction between style and other factors such as the speaker's dialect background and language attitudes in the currently very fluid sociological relationship between Hong Kong Cantonese, the mainland standard of Canton City, Putonghua, etc. (To aid researchers working on this characterization, we recommend that sites that choose to include a phones tier in their C_ToBI transcriptions label the observed segmental alternate on the phones tier rather than regularizing words to the 'underlying' or 'standard' form.)

(2) Segmental alternations in the Cantonese syllables (tones omitted). The symbol 'Ø' represents zero onset.

(*a*)	Not lexically conditioned			
	Onset:	n ~ l	lei ~ nei	you
		ng ~ Ø	ngo ~ o	I
		gw/kw ~ g/k *(before the single vowel /o/)*	gwong ~ gong	bright
	Coda:	n/t ~ ng/k	tek ~ tet	to kick
			hon ~ hong	sweat
(*b*)	Lexically conditioned			
	Onset:	k ~ h	keoi ~ heoi	s/he
		k ~ g	kau ~ gau	to buy
		c ~ j	caai ~ jaai	to step on
		t ~ d	tit (daa) ~ dit (daa)	traditional Chinese physiotherapy
	Coda:	ng ~ k	kwong ~ kwok	to widen
	Vowel:	i ~ e	ling ~ leng	to receive
		a ~ aa	cak ~ caak	to test
		e ~ u	mei (lik) ~ mui (lik)	charm

The examples in (3)–(6) illustrate the tonal alternations, with the examples in (3) showing cases which had previously been analysed in terms of derivational processes resulting in *bianyin* 'changed tone' (Chao 1947) or 'modified tones' (Whitaker 1955–56). That is, these are forms where speakers of some dialects (and classically educated Hong Kong speakers who command literary Cantonese) can still see a more or less transparent relationship between a verb and a derived noun, as in (3*a*), or between a semantically neutral form and a derived form carrying some extra connotation such as familiarity, diminution, derogation, etc. The derived form in each case has a high rising (/35/) tone or a high level (/55/) tone. Much has been written on changed tones since Wong (1941) and the classic articles of Chao (1947), *Cantonese Primer*, and Whitaker (1955–56) (e.g. Wong, S.-L. 1941; Yue-Hashimoto 1972; Wong, M. 1982; Matthews and Yip 1994; Bauer and Benedict 1997), but an adequate description of these alternations today must take into account the changed status of literary Cantonese. That is, many younger Hong Kong Cantonese speakers may not identify the lexical stratum of literary readings for characters which gives classically educated speakers access to the historically 'underlying' form in the alternations in (3).

The loss of a special status for literary Cantonese also affects the analysis of the examples in (4)–(6). Some of the alternations here arise from dialect borrowing of the literary Cantonese form into the spoken language—hence the relationship to casual versus formal style. Educated speakers who command literary Chinese have a clear sense of the different lexical strata, and think of the literary form as having the 'underlying' or 'true' lexical tone and the alternate form as having a 'changed tone' here as well as in (3). Younger speakers, on the other hand, may not distinguish this type of dialect mixture from any other source of alternate forms. In addition to dialectal contact, sources may include collapsing of stylistic differences, tonal dissimilation, and reinterpretation of 'nonce' tones that might have arisen from connected speech effects such as tonal coarticulation. (In keeping with the above recommendation for a phones tier, the C_ToBI convention is to label the observed tonal alternate on the tones tier rather than regularizing to the 'underlying' form—see Section 10.3.1(iii).)

(3) Examples of tone change in Cantonese.

(*a*)	paak3 → 35	to slap → racquet	(nominalization)
(*b*)	daai22 → 55	big → small	(diminution)
(*c*)	wong21 → 35	yellow/surname → egg yolk	(semantic derivation)

(4) Examples of tonal alternations in monosyllabic lexical items.

(*a*)	guk2 ~ 35	bureau
(*b*)	faan23 ~ 22	to over flood
(*c*)	wui23 ~ 33	will (aux. verb)
(*d*)	go35 ~ 55 (go33)	that (one)

(5) Examples of tone alternation in only one of two related forms.

(*a*)	jyu35 / jyu21 pin35	fish / fish fillet
(*b*)	juk35 / juk2 aak35	jade / jade bracelet

(6) Examples of tonal alternations in reduplicative forms.

(*a*)	zim22 *zim22*∼*35*	gradually
(*b*)	kam21 *kam21*∼*35* ceng55	hastily

However, whatever was their origin, none of these tonal alternations in modern Hong Kong Cantonese marks a level of prosodic grouping. That is, they are quite unlike the tonal alternations that define the sandhi group in Xiamen/Taiwanese (Chen 1987; Peng 1997), Fuzhou (Chan 1985), or Shanghainese (Jin 1985; Selkirk and Shen 1990). Thus, there is no obvious basis for prosodic grouping of Cantonese syllables other than the phenomenon of fusion that we have analysed in terms of the foot, and the boundary tones and other phrase final effects that mark the edges of intonation phrases.

10.3. THE CANTONESE TONES AND BREAK INDICES (C_ToBI): ANNOTATION CONVENTIONS

As in all other systems that use the ToBI framework, C_ToBI requires that the transcription include the following three components: (i) a recorded *audio* signal; (ii) a *fundamental frequency* trace for the utterance; and (iii) a set of *symbolic labels* time-linked to the audio and Fo signals. The labelling platform that was used to make the C_ToBI transcriptions in the illustrations in this paper is xwaves and its associated xlabel function. However, any similar labelling platform that can provide the three basic components of the standard ToBI framework transcription may also be used. In developing C_ToBI, we have been careful to define the levels of transcription and inventory of symbolic levels in ways that allow us also to tap the more richly hierarchical field structures of Emu, a speech database labelling platform developed in the Speech Hearing and Language Research Centre at Macquarie University (Cassidy and Harrington 2001).

10.3.1. *Levels of transcription and symbolic labels for C_ToBI*

The C_ToBI system currently specifies the six levels of transcription listed to the right of the xlabel windows displayed in Figure 10.2. These are for tagging

(1) *tones*, (2) *break indices*, (3) any polysyllabic *foot*, (4) *syllables*, (5) *words*, and (6) *miscellaneous* phenomena such as an interval of coughing or an abrupt cut-off of phonation prior to repair of a speech error. The tones tier, words tier, break indices tier, and miscellaneous tier correspond to the original ToBI tiers of the same name. In C_ToBI, however, the words tier is not obligatory (see Section 10.3.1(i)). In the following subsections, the six tiers in C_ToBI are described in turn: words tier, syllables tier, tones tier, break indices tier, foot tier, and miscellaneous tier. This section concludes with suggestions for future tiers.

(i) *The words tier*: in the original ToBI system, the words tier is a word-by-word orthographic transcription. In keeping with the monosyllabic flavour of Cantonese (see Section 10.2.2), the words tier in C_ToBI provides a syllable-by-syllable transcript of the utterance in the native orthography of the language. That is, every syllable is labelled with its corresponding Chinese character, except for syllables that have no native written form, which are instead transcribed in the roman alphabet (see Section 10.3.1(ii)). This tier is not obligatory, because it is not functional on labelling platforms that do not allow Chinese character input using Big5, GB, Unicode, or some other comparable standard encoding system. Chinese character input is currently possible using the Emu labeller in Chinese Windows. In an Emu-based transcription, elements on this level are defined as 'segments' (i.e. tags for discrete intervals with both a beginning and end) rather than as 'events' (tags which are associated with only one time stamp).

(ii) *The syllables tier*: this tier provides an alphabetic transliteration for every element in the words tier. In addition, intervals of silence are flagged with the label <SIL> . The alphabetic transliteration scheme we currently use is the Jyutping Romanization Scheme (1993) developed by the Linguistic Society of Hong Kong (see Appendix I). Unlike the words tier, the syllables tier is obligatory. It serves the function of the words tier for sites that do not have a way to input and/or read Chinese characters. A primary motivation for having a words tier is to allow users of a database to efficiently search for, say, all instances of a given lexical item. Since the syllables tier stands in for the words tier at sites without Chinese character input, we recommend that labels on this tier be made uniform across a speech database. One goal for the C_ToBI labelling community, therefore, should be to develop a freely available communal online character dictionary that regularizes the segmental and tonal alternations described in Section 10.2.5 to a single 'dictionary' form for that morpheme.

(iii) *The tones tier*: the tones tier tags the lexical tones and the boundary tones of an utterance (see Section 10.2.3). It transcribes the lexical tones

Table 10.1 C_ToBI transcription of lexical tones

		Non-checked syllable	Checked syllable
Level tones:	High level	55	5
	Mid level	33	3
	Low level	22	2
Rising tones:	High rising	335	35
	Low rising	223	—
Falling tones:	Low falling	221	21
	High falling*	553*	—

* For those speakers who make a distinction between /55/ and /53/.

syllable by syllable using the traditional Chao-number descriptions with some modification. Specifically, dynamic tones on syllables with all-sonorant rhymes are distinguished from dynamic tones on checked syllables by doubling the first number in the transcription of the non-checked syllables. Thus, /35/ is transcribed as /335/, and /21/ as /221/ in non-checked syllables, to distinguish them from /35/ and /21/ on checked syllables. (Although /23/ and /53/ do not occur on checked syllables in Hong Kong Cantonese, we double the first digit here as well, to make the transcription of non-checked syllables uniform.) The inventory of lexical tones in Hong Kong Cantonese is given in Table 10.1. This list includes two contour tones on checked syllables that are not usually considered part of the inventory of lexical tones in the older linguistics literature—namely, /35/, which is traditionally analysed as a 'changed tone' as in example (3*a*) above, and /21/, which surfaces occasionally on onomatopoeic words. Here, we use 'lexical tones' as a cover term for the full set of tones that contrast monosyllabic morphemes in Hong Kong Cantonese, ignoring intuitions of classically educated speakers who command literary Cantonese.

The tags for boundary tones differ from those for lexical tones. We use 'H' and 'L' and the '%' diacritic from Pierrehumbert (1980), as in the ToBI framework conventions for other languages. In our preliminary analysis, Hong Kong Cantonese has an inventory of six different phrase-final boundary types. Table 10.2 lists the labels for these types and identifies a figure illustrating it.[3] Note that some of these types are distinguished by accompanying durational and/or voice quality effects—for example, 'H%' versus 'H:%' in Figures 10.2 versus 10.7, or the '-%' that marks truncation at phrase-end. These labels reflect our preliminary understanding of the 'surface' features that contrast the boundary types; they are not meant as a

[3] Pragmatic functions of the phrase-final boundary types are suggested in the captions in the figures.

TABLE 10.2 The inventory of boundary tones in C_ToBI

Tone types	Descriptions	Examples
L%	fall from the final lexical tone	Figure 10.3
H%	rise from the final lexical tone	Figure 10.2
H:%	rise from the final lexical tone, with a short plateau at the very end of the rise; incredulity reading accompanied	Figure 10.7
HL%	final rise and then fall from the final lexical tone	Figure 10.4
%	phrase-end with no extra tone	Figure 10.5
-%	truncated rise of the final lexical tone	Figure 10.6
%fi	frame-initial boundary used to mark the initial particle in phrase-framing particle pairs such as '*mat5 ... me55? (rhetorical question)*'	—

definitive phonological analysis. The phrase-initial boundary '%fi' differs in function from the other boundary tags. It identifies the initial syllables of 'pragmatic idioms' that involve the choice of both a phrase-final boundary effect and a pair of particles that 'frame' the construction. Identifying the beginning of the idiom as well as the end will allow us to investigate other potential prosodic markers for these idioms. For example, it is possible that there is a separate pitch range specified for the idiom as a whole.

The phrase-final intonational patterns described in Table 10.2 can occur on a sentence that does not have a final particle, and stand alone to indicate the pragmatic relationship between the sentence and its discourse context. They can also occur with (at least some of) the final particles to produce more complex pragmatic effects, combining the meaning of the intonational pattern with that of the particle, or particle sequence. As noted above, these effects include both tonal phenomena (an 'extra' boundary tone can be added after the lexical tone of the last syllable), and rhythmic/voice-quality phenomena (elongation versus truncation of the last syllable, associated at least sometimes with a contrast between breathy versus creaky or checked voice source). Figures 10.8(a) and 10.8(b) give a minimal pair of examples ending with the final particle *ge335*. The speaker in Figure 10.8(a) asks a question with the question particle *ge335*. Notice that no extra tone is added at the end of the particle, so that it is the particle alone that causes it to be interpreted as a question. By adding the L% at the end of the particle in Figure 10.8(b), the speaker implies further that the answer of the question is already known, as in an English tag question with a falling H* L- L% tune.

When labelling using the xwaves/xlabel platform, a lexical tone label is aligned with the associated words and syllables tier labels, except when the last lexical tone in a phrase needs to precede a boundary. In that case, sites can choose to specify that the lexical tone label be placed just before the boundary tone, or place it earlier—for example, at an appropriate inflection point in the Fo contour. Translating this practice into the Emu labelling system, the tones tier could be treated as an independent stream of 'segments'—in which case the last lexical tone and the boundary tone will arbitrarily divide the last syllable into two intervals. Another possible conversion schema for going between the two labelling platforms treats the tones tier of the xwaves/xlabel platform as a flattened projection of tone labels for two different levels of the prosodic hierarchy. That is, in an Emu labelling template the lexical tones could be specified as one of several independent label fields for syllables, and boundary tones, similarly, would be a label field for intonational phrases. This takes advantage of the more richly hierarchical database structures available with Emu, to encode the different tonal domains directly into the representation of the prosodic hierarchy.

(iv) *The break indices tier*: the set of break indices in the C_ToBI system encodes the levels of grouping identified in Sections 10.2.2 to 10.2.4—namely the syllable, the foot, and the intonational phrase—and the conventions for annotating break indices that were developed in association with the foot tier, the syllables tier, and the tones tier. That is, because of the relationship between 'words' and syllables in the metrical structure of Cantonese (see Sections 10.2.2 and 10.2.2–10.2.4), we can think of the break indices in terms of inter-syllabic juncture, and a break index must be aligned at the end of every label on the syllables tier. Table 10.3 describes the inventory of break labels and diacritics.

Break index 0 marks the weakest disjuncture, for syllable boundaries that are foot internal—i.e. the juncture between syllables in a fusion form, which we now analyse as foot boundary erasure (cf. Sections 10.2.4 and 10.3.1(v)).

Table 10.3 Levels of disjuncture and other diacritics used in the C_ToBI system

Break index/diacritics	Descriptions
0	foot internal syllable boundary
1	end of a syllable that is also end of foot
2	intonation phrase end
1-	uncertainty between 0 and 1
2-	uncertainty between 1 and 2
c	an abrupt, disfluent cut-off of phonation
p	prolongation at a disfluency ('hesitation pause')

Identification of the syllable juncture in a fusion form may not be easy. (Of course, this is a problem for the words and syllables tier labels as well as for the break indices tier.) The fact that syllable fusion in general does not override the lexical tone helps segmentation in cases such as *jai5-22* (fusion form of *jat5 hai22* 'or'). Here, the labels can be placed at the F0 inflection point. However, in cases where the sequence of lexical tones for the fused syllables does not provide an obvious inflection (e.g., in a succession of two identical level tones) and the spectrographic display does not give much clue either (e.g. *zai22-22* as the fusion form of *zau22 hai22* 'is', transcribed on the foot tier as *t+sai22-22*) the '0' can only be placed arbitrarily around the midpoint of the vowel. Break index 1 then marks 'ordinary' syllable juncture, which is also the end of a foot. Break index 2, marking intonation phrase break, is obligatory when a boundary tone is placed on the tones tier.

As in other ToBI framework conventions, these break index numbers can be modified by the '-' uncertainty diacritic. Thus, '1-' marks uncertainty about whether sufficient segmental weakening has occurred to 'erase' the foot boundary, and '2-' marks uncertainty about whether there is an intonational phrase boundary.

The last two tags are modelled after the AmEng_ToBI break-index diacritics for disfluent cut-off or prolongation. The label 'c' marks abrupt cut-off of the syllable, which may or may not be accompanied by a glottal stop, before the sentence continues. The lexical tone may or may not be truncated. The label 'p' marks prolongation of syllables internal to the intonational phrase arising from lexical search and other kinds of hesitation, ground-holding, etc. The diacritic 'p' is not to be used when perceived emphasis is marked for certain syllable(s) on the foot tier. In that case, syllable elongation is a means for focal emphasis, and is marked by '*' on the foot tier (see next section).

In the current C_ToBI annotation conventions, we have defined these as stand-alone labels rather than as diacritics for the break index numbers. Given the dense syntagmatic distribution of tones in Cantonese and our still very incomplete understanding of syllable fusion, it is impossible to make a principled distinction between '1c' and '2c' or between '1c' and '0c' at this time. Also, while there is nothing in theory to preclude a distinction between '1p' and '0p', we have not seen any examples to date of prolongation of the segmental material for the non-final syllable within a fused form.

In translating from xlabel tiers into an Emu labelling system, break indices could be treated as an independent stream of events. Alternatively, the break indices tier could be replaced by an explicit hierarchy of syllables nested within feet, which are in turn nested within intonational phrases (see Section 10.3.1(iii)). The latter move would emphasize the hierarchical relationship

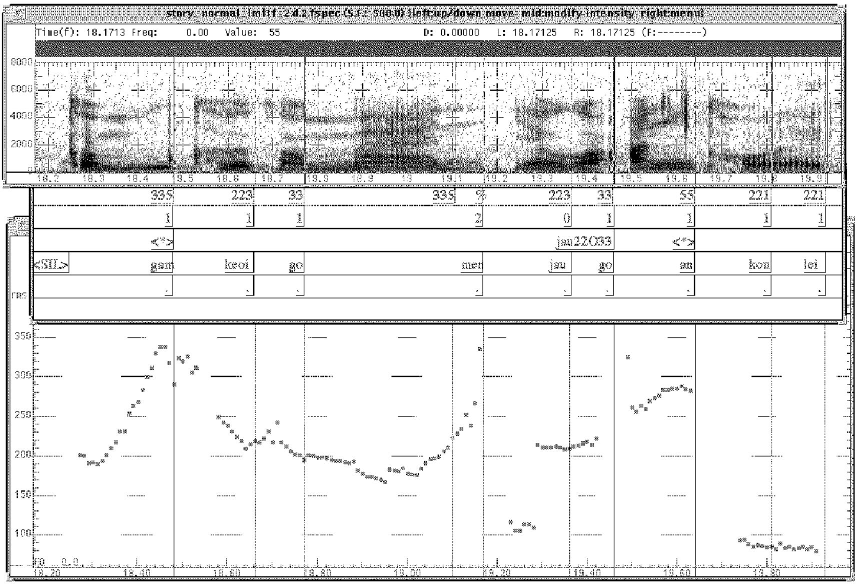


Figure 10.5 F0 contour of the prosodic phrase '*Gam335 keoi223 go33 men335 (Then, his name)*', with no extra tone added after the final syllable *men335*. Notice that lengthening can also occur without adding any extra tones at the boundary (figure transcribed in C_ToBI).

among the different metrical units that the different break index values mark, and would allow the user to take better advantage of Emu's hierarchical querying language.

(v) *The foot tier*: the foot tier is used primarily to tag fusion forms, which we analyse as erasure of the foot boundary. We transcribe these feet with a phonetic transcription of the resulting fused form. (See Appendix I for our set of phonetic symbols, i.e., 'phones' symbols.[4]) The dash notation in the transcription of the form *la221-22* in Figure 10.4 indicates the concatenation of lexical tones in a fusion form. We do not transcribe duration of lexical tones in a fusion form, since this can be measured in the acoustic signal; but truncated rise, as well as falling tone undershoot will be captured. Hence, truncated /23/ and undershoot of /21/ (or *223* and *221* respectively in the C_ToBI transcription of lexical tones; see Table 10.1) will both be transcribed as *22* in a fusion form. (See Figures 10.5, 10.7 and 10.8(a) for examples of truncated rise and Figure 10.8(b) for examples of falling tone undershoot.)

[4] The 'phones' symbols, being all ASCII symbols, were devised with a view to facilitating information exchange (across platforms, operating systems, etc.), database queries, and input efficiency.

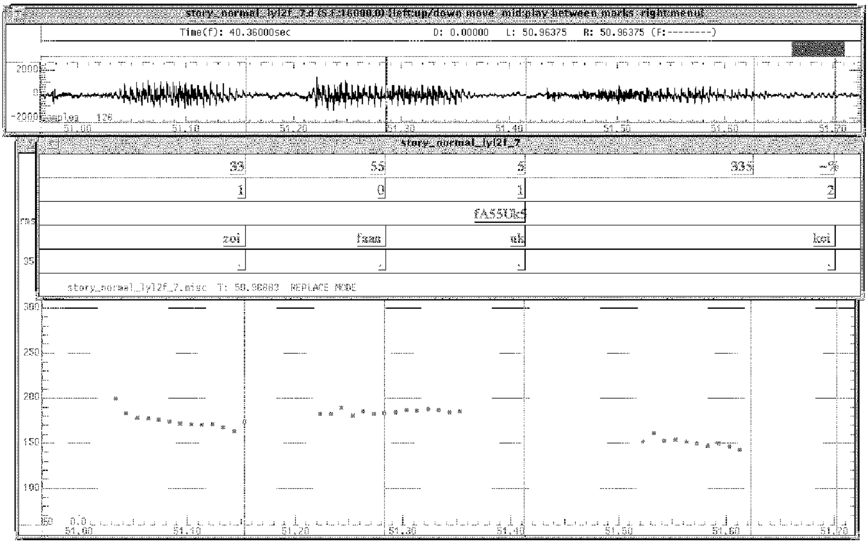


FIGURE 10.6 Fo contour of the utterance '*Zoi33 faan55 uk5 kei335 (And then go home)*', with truncated rise of the final syllable *kei335* (figure transcribed in C_ToBI).

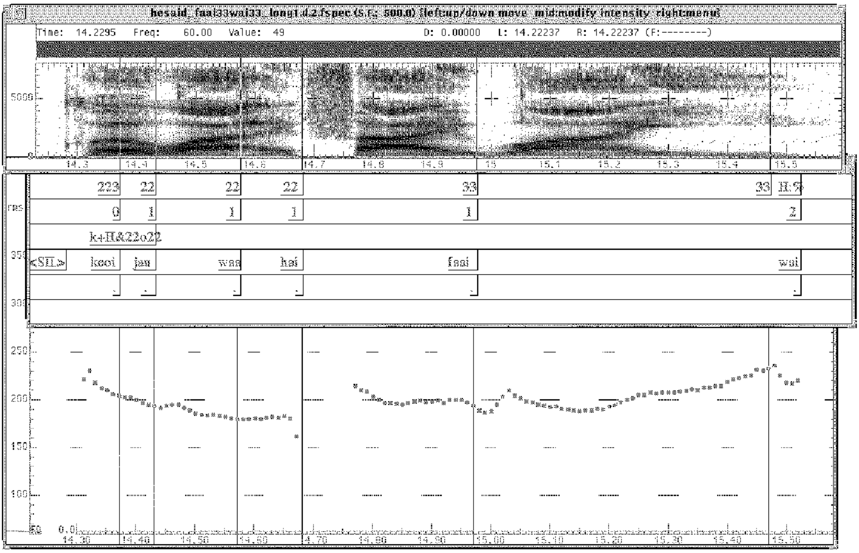


FIGURE 10.7 Fo contour of the utterance '*Keoi223 jau22 waa22 hai22 faai33 wai33? (She then said it was (the word) "pleasant"?!)*', with the pragmatic boundary tone H:% attached to the final tone. The speaker asks with incredulity (figure transcribed in C_ToBI).

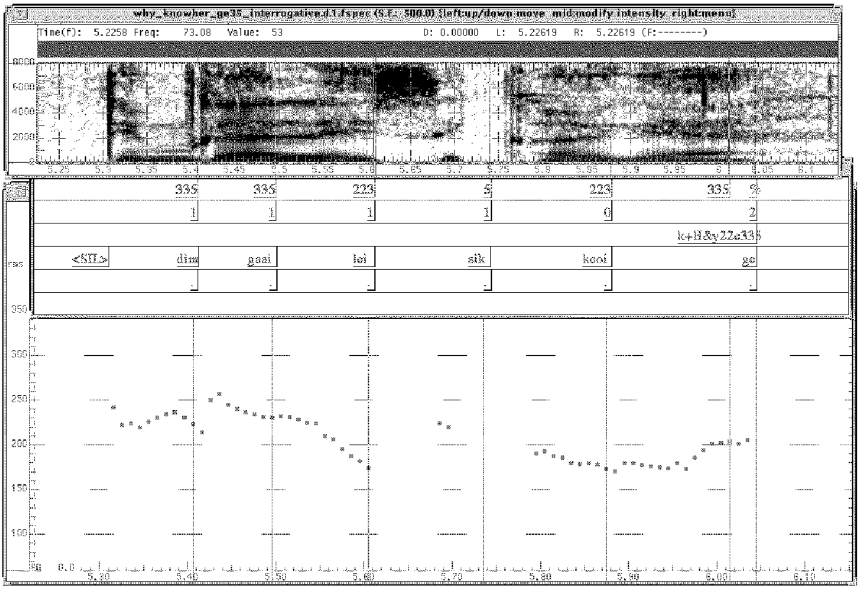


FIGURE 10.8(a) Fo contour of the utterance '*Dim335 gaai335 lei223 sik5 keoi223 ge335[final particle]? (How come you know him/her?)*'. A neutral information-seeking question, where *ge335* is a question particle (figure transcribed in C_ToBI).

We also use the foot tier to tag emphasized syllables and phrases—i.e. monosyllabic 'words' or longer sequences that are set off by an expanded pitch range, longer duration, clearer articulation, and other hallmarks of focal prominence (or phrase-level 'stress'). Understanding the relationship between the domain of these effects and the prosodic hierarchy of syllable, foot, and intonation phrase is an important issue both for the ToBI framework and for phonological theory in general. However, there has been no systematic research on this question for any variety of Cantonese. Thus, our (preliminary) decision to tag these effects on the foot tier is quite arbitrary, and should not be interpreted as a claim that the foot of Cantonese is to be identified with the foot in 'stress' languages such as English and Mandarin. The mark for emphasis is a '*' and the grammar is similar to that on the miscellaneous tier (see Section 10.3.1(vi)). That is, there are paired labels '< *' and '* >' for the beginning and end of an emphasized stretch that is longer than a syllable, and the unitary label '< * >' for a single emphasized syllable. When the first syllable of a fused form is emphasized, we mark it with '< * >' in front of the phonetic transcription. An example is given in Figure 10.5.

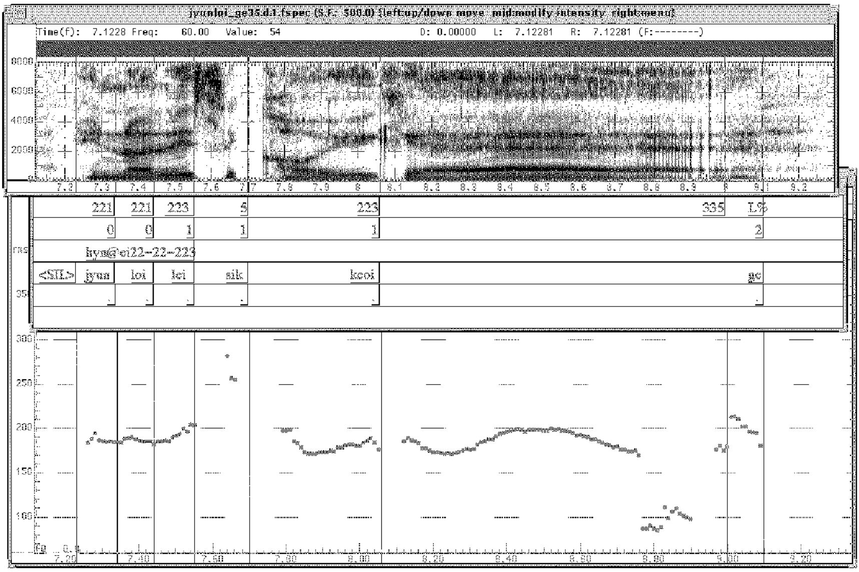


FIGURE 10.8(b) F0 contour of the utterance '*Jyun221 loi221 lei223 sik5 keoi223 ge335[final particle] (So I see you know him/her)*'. With L% added after the final particle *ge335*, the speaker indicates that the answer to the question has been made known to her. The question particle is now said in combination with assertiveness expressed by the L tone pragmatic morpheme (figure transcribed in C_ToBI). Notice how creaky the voice quality of the L% can be (such creakiness may cause pitch tracking failure and result in pitch-doubling as can be seen at the phrase-end of this utterance).

In translating this tier into a hierarchical Emu labelling template (as suggested in earlier sections), the fusion form transcriptions could be treated as another field of tags for the foot tier segments. The emphasis tags, however, would need to be treated as a separate stream of segment labels, since the relationship between the domain of these effects and the prosodic hierarchy is not yet known. This highlights one of the principal advantages of the looser structure of the original ToBI framework. While the hierarchy of break indices is amenable to interpretation in terms of the prosodic hierarchies proposed in work such as Nespor and Vogel (1986), Selkirk and Shen (1990), or Keating *et al.* (2003), the interpretation is not enforced. That is, the architecture of independent parallel tiers does not require that the domain of an effect that is tagged on another tier be identified with any particular break index level unless there is sufficient theory-external evidence for it.

(vi) *The miscellaneous tier*: as in other ToBI systems, this tier is used to mark breaths, coughs, disfluencies, etc., using labels with '< ' and ' >' to mark roughly the beginning and the end points of the labelled effect—for example, 'laugh < ' and 'laugh >' for an interval where speech is interrupted by laughter. Finer specifications may be devised according to site-specific needs.

(vii) *Other tiers*: we recommend that at sites where there are resources to do a phoneme-by-phoneme segmental labelling, there should also be a phones tier. Appendix I lists our proposed inventory of phones symbols. (Both the IPA and Jyutping symbols are given.) We also recommend the eventual development of a sociolinguistic variables tier and a codes tier, given the facts in Section 10.2.5. The detailed working out of these other tiers should be modelled on annotation conventions in other communities which have adopted the ToBI framework where development of such tiers is well underway.

10.4. DISCUSSION AND FUTURE WORK

The conventions for transcribing spoken Cantonese corpora developed thus far encode what we know already about Cantonese prosody, as well as our 'best guess' hypotheses about phenomena that need to be further tested before the labelling conventions can be 'frozen' for large-scale database development. (A summary of the C_ToBI labels is given in Appendix II.) In this section, we summarize the hypotheses that seem most solidly grounded in current knowledge, and the facts on which they are based, before listing further questions for research.

We know, for instance, that boundary tones come at the edges of prosodic constituents larger than the word. In some speech styles, these edges can be marked by a very salient elongation of the final syllable's rhyme, and these may involve utterances with or without final particles. Particles in sequence can easily exceed a second in duration (Chan 1998), a phenomenon that seems to be particularly salient in Cantonese. A deliberate truncation can also occur here, with a definite pragmatic effect that seems to interact with sentence particle meaning in the same way that boundary tones do. For example, the final particle *zek5* in this variant—with checked syllable and shorter duration—is a stronger way to voice a complaint than the corresponding open syllable variant, *ze55*, which can be elongated for different pragmatic effects (Chan 1996 and sources cited therein; Fung 2000). Further work is needed to explore the precise nature of these interactions, and what stylistic and other constraints there are on both truncation and prolongation.

We also know that fused forms exist, and that they are more common at faster tempo (see Wong, P. W.-Y. 1996). Li (1986) also notes style of speech, frequency of use, and segmental make-up as possible factors for contraction/fusion. There may also be speech register and regional dialectal influences. Further research is needed to understand different aspects of this phenomenon, including the extent to which it occurs across the different varieties of Cantonese. Besides Hong Kong Cantonese, fusion has also been reported for Toisan (Taishan) Cantonese (Kong 1984), for example. We currently analyse fusion as an erasure of the foot boundary, because the syllable edge becomes blurred without a necessary reduction in syllable count in all cases (contra an earlier analysis in Wong, P. W.-Y. 1996). Further corpus work combined with psycholinguistic experiments should be helpful in testing these two competing analyses.

Also, potential constraints from emphasis/stress on the potential for fusion forms need to be explored. Our native speaker intuition is that if just one of the syllables in a fused form is emphasized, it has to be the first. If this turns out to be true in the databases that we develop with this annotation tool, then we will have evidence of the possible beginnings of a more Mandarin-like stress system, albeit realized in segmental lenition without tone deletion.

Prosodically annotated speech corpora of Cantonese are essential for answering these and other questions. We hope that results of corpus studies will feed back into the development of the labelling system, appropriately tuning the current set of labels, and rendering C_ToBI an efficient tool for using annotated speech corpora for diverse research purposes and technological applications.

APPENDIX I

A correspondence table showing the Jyutping Romanization Scheme (1993) for Cantonese < http://www.hku.hk/linguist/lshk/ >, the phonetic symbols (the 'phones' symbols) that we use in transcribing fusion forms on the foot tier (and for sites having a phones tier), and the corresponding IPA symbols (length distinction of nuclear vowels are indicated (cf. Lee 1999; Zee 1995, 1999*a*, 1999*c*)).

Consonants			**Vowels**			**Diphthongs**		
IPA	**Phones**	**Jyutping**	**IPA**	**Phones**	**Jyutping**	**IPA**	**Phones**	**Jyutping**
pʰ	p+H	p	iː	i	i	uːi	ui	ui
p	p	b	yː	y	yu	ei	ei	ei
tʰ	t+H	t	uː	u	u	ɔːi	Oi	oi
t	t	d	ɪ	I	i	ɐi	ai	ai

k^{h}	k+H	k	ʊ	U	u	aːi	Ai	aai
k	k	g	ɛː	E	e	iːu	iu	iu
m	m	m	œː	R	oe	ɛːu	Eu	eu
n	n	n	θ	&	eo	ou	ou	ou
ŋ	N	ng	ɔː	O	o	ɐu	au	au
f	f	f	ɐ	a	a	aːu	Au	aau
s	s	s	aː	A	aa	θy	&y	eoi
h	h	h	**Syllabic**	**nasals**		**Others**		
w	w	w	**IPA**	**Phones**	**Jyutping**	**IPA**	**Phones**	**Jyutping**
j	j	j	m̩	M	m	ə	@	—
l	l	l	ŋ	N	ng	æ	X	—
ts^{h}	t+s+H	c				ɥ̩	W	—
ts	t+s	z				ts^{j}	T+S	z
k^{wh}	kw+H	kw				ts^{hj}	T+S+H	c
k^{w}	kw	gw				ʃ	S	s
						ʔ	q	—
						∼	∼*	—

* The symbol '∼' is to be added after the segment that is nasalized.

APPENDIX II

A summary of the C_ToBI labels is presented below. Pragmatic functions of boundary tones are given here for illustrative purposes only, since boundary tones may interact with final particles (occurring singly or in combinations) to produce a complex set of pragmatic effects.

L% *IP-final fall*: marked on the right edge of intonational phrases after the final lexical tone in declarative statements.

H% *IP-final rise*: marked on the right edge of intonational phrases after the final lexical tone signalling confirmation-seeking questions (with possible surprise).

H:% *IP-final rise with short plateau at the very end of the rise*: marked on the right edge of intonational phrases after the final lexical tone in questions expressing incredulity.

HL% *IP-final rise-fall*: marked on the right edge of intonational phrases after the final lexical tone expressing connotation of 'discovery.'

% *IP-final*: phrase-end with no extra tone.

-% *IP-final truncated rise*: marked on the right edge of intonational phrases with truncated rise of the final lexical tone.

%fi	*IP frame-initial*: frame-initial boundary used to mark the initial particle in phrase-framing of particle pairs such as "*mat5 . . . . me55 ? (rhetorical question).*"
0	*Break index: weak disjuncture.* indicates foot-internal syllable boundaries of fused forms.
1	*Break index: medium disjuncture.* indicates the end of a syllable that is also the end of a foot; for 'ordinary' syllable juncture.
2	*Break index: strong disjuncture.* indicates the end of an intonational phrase
1-	*Break index uncertainty between 0 and 1.*
2-	*Break index uncertainty between 1 and 2.*
c	*Cutoff*: indicates an abrupt, disfluent cutoff of phonation.
p	*Prolongation*: indicates prolongation at a disfluency ("hesitation pause").

REFERENCES

Arvaniti, A., and Baltazani, M. (this volume Ch. 4), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Corpora'.

Bauer, R. S. (1979), 'Alveolarization in Cantonese: A Case of Lexical Diffusion', *Journal of Chinese Linguistics*, 7: 132–41.

——, and Benedict, P. K. (1997), *Modern Cantonese Phonology. Trends in Linguistics: Studies and Monographs 102* (Berlin and New York: Mouton de Gruyter).

Beckman, M. E., and Elam, G. A. (1997), 'Guidelines for ToBI Labelling', version 3, Department of Linguistics (Ohio State University), accessible from the World Wide Web: < ftp://julius.ling.ohio-state.edu/pub/TOBI/DOCS/labelling_guide_v3.ASCII > (short extract of the guidelines is accessible at: < http://www.ling.ohio-state.edu/phonetics/E_ToBI/ >) [Accessed 30 July 2000].

——, and Jun, S.-A. (1996), 'K-ToBI (Korean ToBI) Labeling Convention, version 2.1', Department of Linguistics (Ohio State University) and Department of Linguistics (University of California, Los Angeles), accessible from the World Wide Web: < http://www.humnet.ucla.edu/humnet/linguistics/faciliti/facilities/k_tobi.html > [Accessed 30 July 2000].

Bourgerie, D. S. (1990), 'A Quantitative Study of Sociolinguistic Variation in Cantonese', Ph.D. dissertation (Ohio State University).

Bruce, G. (this volume Ch. 15). 'Intonational Prominence in Varieties of Swedish Revisited'.

Cassidy, S., and Harrington, J. (2001), 'Multi-level Annotation in the Emu Speech Database Management System', *Speech Communication*, 33/1–2: 61–77.

Chan, M. K.-M. (1985), 'Fuzhou Phonology: A Non-linear Analysis of Tone and Stress', Ph.D. dissertation (University of Washington).

Chan, M. K.-M. (1996), 'Gender-marked Speech in Cantonese: The Case of Sentence-final Particles *je* and *jek*', *Studies in the Linguistic Sciences*, 26/1–2: 1–38.

—— (1998), 'Review of Matthew, S. and Yip, V., 1994', *Journal of the Chinese Language Teachers Association*, 33/3: 97–106.

——, Wong, W.-Y. P., and Beckman, M. E. (1998), 'Utterance-final Phenomena in Cantonese', presentation at the Pragmatics study group at the Linguistics Department, Ohio State University, 5 November.

Chao, Y.-R. (1947), *Cantonese Primer* (Cambridge, MA: Harvard University Press).

Chen, M. Y. (1987), 'The Syntax of Xiamen Tone Sandhi', *Phonology Yearbook*, 4: 109–49.

Cheung, K.-H. (1986), 'The Phonology of Present-day Cantonese', Ph.D. dissertation (University College London).

Duanmu, S. (1990), 'A Formal Study of Syllable, Tone, Stress and Domain in Chinese Languages', Ph.D. dissertation (MIT), in *MIT Working Papers in Linguistics.*

Fung, R. S.-Y. (2000), 'Final Particles in Standard Cantonese: Semantic Extension and Pragmatic Inference', Ph.D. dissertation (Ohio State University).

Grice, M., Baumann, S., and Benzmüller, R. (this volume Ch. 3), 'German Intonation in Autosegmental-Metrical Phonology'.

Grimes, B. F. (1996), (eds.) *Ethnologue: Languages of the World* (13th edn., Dallas, Texas: Summer Institute of Linguistics), accessible from the World Wide Web: < http://www.sil.org/ethnologue/ > [Accessed 25 January 2000].

Hong Kong Census and Statistics Department (1996), *Population By-census* (Hong Kong Government).

—— (2001), *Population Census* (Hong Kong Special Administrative Region), accessible from the World Wide Web: Ethnicity and language use— < http://www.info.gov.hk/censtatd/chinese/hkstat/fas/01c/01c_index.html > [Accessed 2 January 2002].

Jin, S.-D. (1985), 'Shanghai Morphotonemics: A Preliminary Study of Tone Sandhi Behavior Across Word Boundaries', thesis (University of Pittsburgh), (Bloomington, IN: Indiana University Linguistics Club).

Jun, S.-A. (this volume Ch. 8), 'Korean Intonational phonology and Prosodic Transcription'.

Keating, P., Cho, T., Fougeron, C., and Hsu, C.-S. (2004), 'Domain-initial Articulatory Strengthening in Four Languages', in J. Local, R. Ogden, and R. Temple, (eds.), *Papers in Laboratory Phonology VI* (Cambridge: Cambridge University Press), 143–61.

Kong, B. (1984), 'Assimilation, Contraction, and Word Fusion in Toisan', ms, University of British Columbia.

Kwok, H. (1984), *Sentence Particles in Cantonese* (University of Hong Kong: Centre of Asian Studies).

Lee, W.-S. (1999), 'A Phonetic Study of the Speech of the Cantonese-Speaking Children in Hong Kong', Ph.D. dissertation (City University of Hong Kong).

Leung, C.-S. (1992), 'A Study of the Utterance Particles in Cantonese as Spoken in Hong Kong', thesis (Hong Kong Polytechnic University).

Li, P. Y.-C. (1986), 'Contraction in Cantonese: A First Probe', ms, University of California, San Diego.

Matthews, S., and Yip, V. (1994), *Cantonese: A Comprehensive Grammar* (London: Routledge).

Nespor, M., and Vogel, I. (1986), *Prosodic Phonology* (Dordrecht: Foris).

Newman, J. (1987), 'The Evolution of a Cantonese Phonotactic Constraint', *Australian Journal of Linguistics*, 7: 43–72.

Peng, S.-H. (1997), 'Production and Perception of Taiwanese Tones in Different Tonal and Prosodic Contexts', *Journal of Phonetics*, 25/3: 371–400.

——, Chan, M. K.-M., Tseng, C.-Y., Huang, T., Lee, O.-J., and Beckman, M. E. (this volume Ch. 9), 'Towards a Pan-Mandarin System for Prosodic Transcription'.

Pierrehumbert, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

Rao, O.-Y., Ouyang, J.-Y., and Zhou, W.-J. (1981), *Guangzhouhua Fangyan Cidian* [Cantonese Dialect Dictionary] (Hong Kong: Commercial Press).

Selkirk, E., and Shen, X. (1990), 'Prosodic Domains in Shanghai Chinese', in S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection* (Chicago: University of Chicago Press).

Silverman, K., Beckman, M. E., Pitrelli, J., Ostendorf, M., Wightman C., Price P., Pierrehumbert, J., and Hirschberg, J. (1992), 'ToBI: A Standard for Labeling English Prosody', in *Proceedings of the 1992 International Conference on Spoken Language Processing* (Banff, Canada), 2: 867–70.

Venditti, J. J. (this volume Ch. 7), 'The J_ToBI model of Japanese intonation'.

Whitaker, K. P. K. (1955–56), 'A Study of the Modified Tones in Spoken Cantonese', *Asia Major: New Series*, 5: 9–36, 184–207.

Wong, M. (1982), 'Tone Change in Cantonese', Ph.D. dissertation (University of Illinois at Urbana-Champaign).

Wong, W. Y. P. (1996), 'Tempo, Processing Rate and Clarity Drive in Hong Kong Cantonese Connected Speech', thesis (Hong Kong Polytechnic University).

—— (2002*a*), 'Cues and Constraints: Production and Perception of Coda Place in Modern Hong Kong Cantonese', unpublished manuscript (First Pre-Candidacy Paper), Department of Linguistics, Ohio State University.

—— (2002*b*), 'Syllable Fusion and Speech Rate in Hong Kong Cantonese', unpublished manuscript (Second Pre-Candidacy paper), Department of Linguistics, Ohio State University.

Wong, S.-L. (1941), *A Chinese Syllabary Pronounced According to the Dialect of Canton* (Hong Kong: Chung Hwa Book Co.).

Yau, S.-C. (1980), 'Sentential Connotations in Cantonese', *Fangyan*, 1: 35–52.

Yeung, H. S.-W. (1980), 'Some Aspects of Phonological Variation in the Cantonese Spoken in Hong Kong', thesis (University of Hong Kong).

Yip, M. (1980), 'The Tonal Phonology of Chinese', Ph.D. dissertation (Massachusetts Institute of Technology) (Bloomington, IN: Indiana University Linguistics Club).

Yip, M. (1988), 'The Obligatory Contour Principle and Phonological Rules: A Loss of Identity', *Linguistic Inquiry*, 19: 65–100.

Yuan, J.-H. ([1960], 1983), *Hanyu Fangyan Gaiyao* [Survey of Chinese Dialects] (Beijing: Wenzi Gaige Chubanshe).

Yue-Hashimoto, O.-K. (1972), *Phonology of Cantonese* (Cambridge, UK: Cambridge University Press).

Zee, E. (1995), 'Temporal Organisation of Syllable Production in Cantonese Chinese', *Proceedings of ICPhS 95* (Stockholm), 3, 250–3.

—— (1999*a*), 'An Acoustical Analysis of the Diphthongs in Cantonese', *Proceedings of ICPhS 99* (San Francisco), 2, 1101–04.

—— (1999*b*), 'Change and Variation in the Syllable-initial and Syllable-final Consonants in Hong Kong Cantonese', *Journal of Chinese Linguistics*, 27/1: 120–67.

—— (1999*c*), 'Resonance Frequency and Vowel Transcription in Cantonese', *Proceedings of the 10 th North American Conference of Chinese Linguistics and the 7th Annual Meeting of the International Association of Chinese Linguistics*, Harvard University, Cambridge, MA.

Zhan, B.-H., and Cheung, Y.-S. (1987), *A Survey of Dialects in the Pearl River Delta, Vol. 1: Comparative Morpheme-Syllabary* (Hong Kong: New Century Publishing House).

11

Intonational Phonology of Chickasaw

Matthew K. Gordon

11.1. INTRODUCTION

This chapter presents a formal model for transcribing the intonational properties of Chickasaw, a Western Muskogean language spoken in south-central Oklahoma by perhaps a few hundred speakers. The proposed model adopts many elements and assumptions of the autosegmental/metrical (AM) framework originally developed to analyse English intonational structure (Pierrehumbert 1980; Silverman *et al.* 1992; Beckman and Hirschberg 1994; Pitrelli *et al.* 1994; Beckman *et al.* (this volume)) and subsequently extended to other languages, such as Japanese (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Campbell and Venditti 1995; Venditti 1995, this volume Ch. 7), Korean (Jun 1993, this volume Ch. 8), French (Jun and Fougeron 1995), and other languages described in this volume. Because its prosodic system differs rather substantially from most other languages analysed within the AM paradigm, Chickasaw is an interesting test case for establishing the ability of the AM framework to model a broad range of

The author owes a great debt of gratitude to Pam Munro for her guidance, advice, and encouragement in this project, and to the Chickasaw speakers (Frankie Alberson, Adeline Brown, Willie Byars, Onita Carnes, Thomas and Lizzie Frazier, Jerry Imotichey, Mary James, William Pettigrew, Eloise Pickens, Lee Fannie Roberts, the late Mary Ella Russell, Thomas Underwood, Jimmie Walker, and Catherine Willmond) who so graciously provided the data reported here. Many thanks also to Sun-Ah Jun for her advice in intonation labelling, to Judith Bishop and Sun-Ah Jun for their insightful comments on earlier drafts of this paper, and to Bruce Hayes, Peter Ladefoged, and Pam Munro for comments and suggestions. Thanks also to audiences at the Intonation Workshop held in conjunction with the 14th International Congress of Phonetic Sciences in San Francisco, as well as to audiences at the 14th International Congress of Phonetic Sciences, at the 2nd WAIL meeting at UCSB, at the 1996 Acoustical Society of America meeting in Hawaii, and at UCLA and UCSB for comments and suggestions. Any remaining errors and shortcomings are my own responsibility. This work was partially funded by an NSF grant SBR 9511118 to Peter Ladefoged and Ian Maddieson, and by the UCLA Linguistics Department.

intonation systems (see also Bishop and Fletcher this volume Ch. 12 for analysis of the intonation of Bininj Gun-wok, another language prosodically quite different from others modelled within the AM framework). Although the Chickasaw data require modifications of certain aspects of the AM systems employed by investigators of other languages, this chapter will show that the AM framework is ultimately well suited to the analysis of Chickasaw intonation.

The structure of this chapter is as follows. Section 11.2 provides an overview of Chickasaw prosody. Section 11.3 focuses more narrowly on Chickasaw intonation and its interactions with other aspects of the prosodic system. Section 11.4 examines the hierarchical prosodic structure of Chickasaw utterances. Section 11.5 proposes an AM model for transcribing Chickasaw intonation. Finally, Section 11.6 summarizes the principal findings and suggests areas for future research in Chickasaw prosody and intonation.

11.2. CHICKASAW PROSODY

Chickasaw has a relatively complex prosodic system combining aspects of a stress system with those of a lexical pitch accent system. Furthermore, a process of rhythmic vowel lengthening contributes an additional layer of durational prominence. Sections 11.2.1–11.2.3 introduce these prosodic features, all of which interact with the intonation system. Rhythmic lengthening is discussed in Section 11.2.1, word-level stress is examined in Section 11.2.2, and the system of morpholexical pitch accents is covered in Section 11.2.3.

11.2.1. *Rhythmic lengthening*

One of the more salient characteristics of Chickasaw prosody is its pattern of rhythmic lengthening whereby the second in a sequence of two vowels in adjacent open syllables is phonetically lengthened (see Munro and Ulrich 1984; Munro and Willmond 1994; Munro 1996, 1999, forthcoming, for discussion). Thus, for example, the underlying string /pisalitok/ 'I looked at it' is realized phonetically as [pisaˑlitok], where the rhythmically lengthened [aˑ] is indicated by a half-length IPA symbol. (Rhythmically lengthened vowels are not differentiated from unlengthened vowels in the orthography.) Rhythmic lengthening does not affect vowels in final position of the morphological word; thus, the word /pisa/ 's/he looks at it' is realized simply as [pisa] without lengthening of the final vowel. Rhythmic lengthening is generally non-neutralizing; lengthened vowels are usually slightly shorter than phonemic long vowels, though certain speakers do neutralize the two length categories for certain vowel qualities (see Gordon *et al.* 2000 for phonetic measurements).

Certain prefixes and suffixes belonging to the morphological word fall outside the domain of rhythmic lengthening. For example, in the word /im-apila-li-tok/ 'I helped him/her for him/her', the dative prefix im- falls outside of the rhythmic lengthening domain which initiates with the second syllable, the first syllable of the root –apila-; the result is lengthening of the third and fifth and not the second and fourth vowels, i.e. [imapiˑlaliˑtok]. There are certain other complications regarding rhythmic lengthening which will not concern us here (see the aforementioned works for discussion).

A basic understanding of rhythmic lengthening is necessary in discussing Chickasaw intonation, since the lengthened vowels behave identically to phonemic long vowels for purposes of nuclear pitch accent placement. Rhythmically lengthened vowels also behave parallel to phonemic long vowels with respect to other prosodic and morphological phenomena in Chickasaw (see Munro and Willmond 1994 for details).

11.2.2. *Stress*

Stress is not phonemic in Chickasaw; rather, the Chickasaw stress system is weight-sensitive, in the sense that certain 'heavy' syllable types preferentially attract stress. Chickasaw observes a three-way hierarchy of weight which manifests itself in both the stress and nuclear pitch accent system (see Section 11.3.2): long and lengthened vowels (CVV) are heaviest, followed by closed syllables (CVC), which, in turn, are heavier than open syllables containing a short vowel (CV). With respect to stress, Munro (1996) observes that the final syllable of a word is prominent, as are closed syllables and syllables containing long vowels (including phonemic long vowels, rhythmically lengthened vowels, and nasalized vowels, all of which are phonetically long). Thus, CVV and CVC are heavier than CV, which is unstressed unless in final position. Primary stress falls on the rightmost long (or lengthened vowel) in a word, indicating that CVV is heavier than CVC. In the absence of any long vowels, the primary stress falls on the final syllable of a word. The words in (1) illustrate word-level stress patterns.

(1) (*a*) ˌbakʃiˈjaˑˌmaʔ 'diaper'
(*b*) aˈboːkoˌʃiʔ 'river'
(*c*) tʃaˌlakˈːiʔ 'Cherokee'

(*d*) ˌokˌfokˈːol ‘type of snail’
(*e*) ˈnaːɬtoˌkaʔ ‘policeman’

Primary stressed syllables are often associated with heightened fundamental frequency and increased intensity and duration; these properties are exaggerated when realized on a long or lengthened vowel. Secondary stress, which falls on CVV and CVC and on final syllables not carrying primary stress, is often associated with increased duration and intensity, though the presence of these properties is inconsistent. Perhaps the most reliable diagnostic for distinguishing between secondary stressed and unstressed syllables is a series of syncope processes affecting light, unstressed syllables, i.e. non-final CV. Munro and Willmond (1994) and Munro (1996) describe a number of these processes, which include deletion of word-medial –li– containing a non-rhythmically lengthened, i.e. unstressed, vowel at the end of a verb stem (2*a*), and syncope of a non-lengthened (unstressed) vowel between a strident and a coronal (2*b*) (see Munro and Willmond 1994 for discussion of other syncope rules).

(2) (*a*) maˈliˑli-ˌtok → [maˈliˑˌtok] ‘S/he ran.’
(*b*) piˈlaˑtʃiˌtok → [piˈlaˑʃˌtok] ‘S/he sent it.’
(tʃ → ʃ / __[+coronal] by regular rule)

11.2.3. *Morpholexical pitch accents*

In Chickasaw, a subset of verbs are marked as carrying a morpholexical pitch accent, phonetically a high tone, on a particular syllable. Unlike in prototypical pitch accent languages like Japanese (Venditti Ch. 7 this volume) and Serbo-Croatian (Godjevac Ch. 6 this volume), verbs carrying one of these morpholexical pitch accents are often (though not always) semantically and phonologically related to a base word from which they are derived, though the precise semantic relationship between the base and the morpholexically accented form is not transparent in many cases; hence the term ‘morpholexical’ pitch accent used here. Morpholexically accented forms, termed verb ‘grades’ by scholars of Chickasaw and related Muskogean languages, convey aspectual information, such as intensification, active and stative changes, habitual action, among other properties. There are various classes of verb grades which differ in their semantic and phonological relationship with the base from which they are derived (see Munro and Willmond 1994: lv–lxii for detailed discussion of Chickasaw grades). Crucially for present purposes, morpholexically pitch accented forms contain one syllable which carries a high tone, the penult in most unsuffixed grade forms, but the antepenult or preantepenult in certain grades. In addition, there are other segmental changes accompanying grade formation, such as gemination, nasalization, or laryngeal insertion, which are irrelevant for purposes of the present discussion. Some examples of morpholexically pitch accented forms and their related bases appear in (3). Morpholexical pitch accents are indicated by a circumflex accent.

		Grade form	*gloss*	*base*	*gloss*
(3)	(*a*)	hîkːiʔja	be standing	hika	stand up
	(*b*)	tʃofânta	be cleaner	tʃofaˑta	be clean
	(*c*)	itːibâkːakliʔtʃi	make a knocking sound	bakaʔtʃi	make a noise with wood
	(*d*)	malîːli	to start, run (car engine)	maliˑli	run
	(*e*)	totʃːîʔna	be three		none

An interesting property of morpholexically pitch accented syllables is that they are always heavy (either CVV or CVC), either because they are closed, as in (3*a–c*) and (3*e*) above or because they contain a long vowel, as in (3*d*).

11.3. INTONATIONAL PHONOLOGY

Three types of phonological tones are relevant for analysing Chickasaw intonation. The first type of tone is the boundary tone which occurs at the right edge of the largest intonational constituent, the Intonational Phrase, or IP (see Section 11.4.1 for discussion of the Intonational Phrase). Boundary tones are discussed in Section 11.3.1. The second category of tones includes the pitch accents, which occur in two varieties: phonological pitch accents and morpholexical pitch accents associated with certain lexically marked syllables, as described in Section 11.2.3. Pitch accents are discussed in Section 11.3.2. Finally, the smallest intonational constituent, the Accentual Phrase, is characterized by a series of phrasal tones linked to certain positions within the Accentual Phrase. Discussion of the Accentual Phrase tones is deferred until Section 11.4.2 (i) after analysis of the prosodic organization of Chickasaw.

11.3.1. *Boundary tones*

Contrary to the dominant cross-linguistic pattern, Chickasaw speakers usually end a statement with a final rise in fundamental frequency. In fact, the

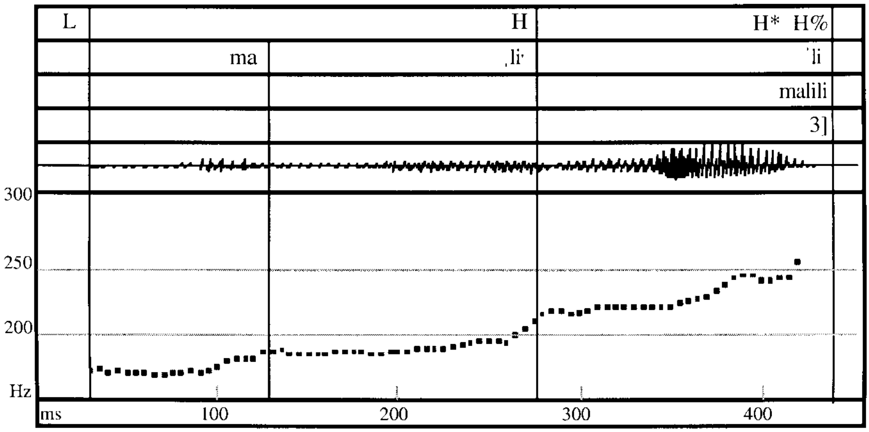


Figure 11.1 Final H% boundary tone in statement IP 'Malili', 'S/he runs' (female speaker). Note that an IPA transcription of the utterance appears in the transcription tier below the tones and above the orthographic transcription (see Section 11.5 for transcription guidelines).

highest fundamental frequency in a Chickasaw statement typically occurs on the final syllable, reflecting either an H% boundary tone or the H* pitch accent on the ultima (see Section 11.3.2) followed by no boundary tone. A statement ending in a H% boundary tone is illustrated in Figure 11.1.

Echo questions characteristically also end in a H% boundary tone, with the principal difference in intonation between an echo question and a statement residing in the overall increase in fundamental frequency characteristic of an echo question. This increase in pitch level is apparent throughout an echo question, such that all low tones and high tones, including boundary tones, are higher in a echo question than in a statement.

Unlike echo questions, both wh- and yes/no- questions in Chickasaw end in a pitch fall commencing immediately after the nuclear pitch accent (see Section 11.3.2 for discussion of nuclear pitch accents) and persisting until the end of the question. The lowest pitch in a question is found at the end, indicating a L% boundary tone. The L% boundary tone is also found in exclamations expressing surprise or disbelief. A question exemplifying the L% boundary tone appears in Figure 11.2.

A L% boundary tone is also found in sentences in which SVO and OVS word orders are employed rather than the more standard SOV order. In SVO and OVS sentences, the first noun and the verb form one Intonational Phrase (see Section 11.4.1 for the Intonational Phrase), whose intonational properties, including boundary tones, mirror those found in SOV sentences.

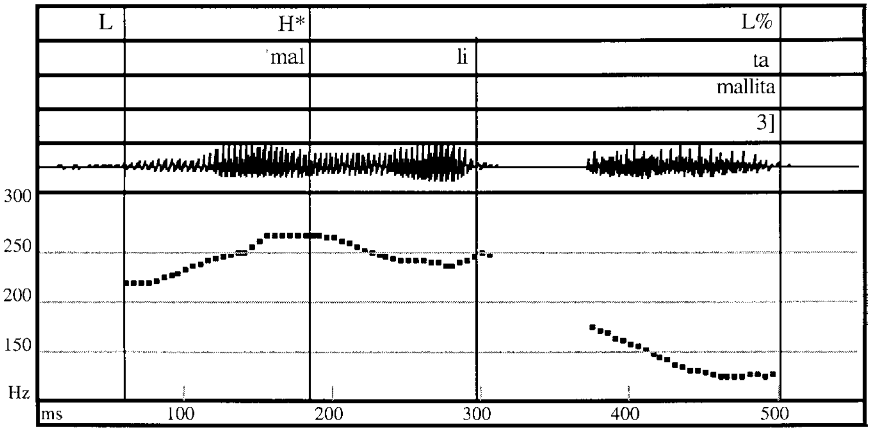


Figure 11.2 Nuclear pitch accent in question IP 'Mallita?' 'Does she/he jump?' (female speaker). Note that the slight fo increase before the /t/ is a segmental effect.

The postverbal noun, on the other hand, constitutes a separate Intonational Phrase which ends in a L% boundary tone, even in statements. The postverbal noun is also characterized by an overall lower and compressed pitch range, meaning that the L% boundary tone at the end of the postverbal noun is the point of lowest fundamental frequency in the utterance. This compression of the pitch range affects the realization of pitch accents within the postverbal noun; they are realized as downstepped accents (see discussion in Section 11.3.2).

A L% boundary tone is also found at the end of non-main clauses, some of which are translated as subordinate clauses in English and others of which are coordinate clauses (see Munro and Willmond 1994 for discussion). In case the non-main clause follows (but not when it precedes) the main clause, the non-main clause is associated with the same overall lowered and compressed pitch range characteristic of postverbal nouns. Figure 11.3 illustrates the intonation associated with a biclausal sentence in which the second clause is a non-main clause. In Figure 11.3, the first clause [maˌliʼˈli] in the utterance [maˌliʼˈli ˌnambiˌlaʼˈmaːt ˈalaˌkãː] 'S/he runs, when there's perfume here' is realized with the H% characteristic of main clauses, while the second clause [ˌnambiˌlaʼˈmaːt ˈalaˌkãː] has a L% final boundary tone and also a reduced and lowered pitch range relative to the main clause.

In addition to the H% and L% boundary tones, a HL% boundary tone is occasionally found in statements and is consistently found in imperatives. The use of the HL% boundary as an alternative to the H% boundary tone in statements appears to be to a large extent a speaker specific matter. Certain

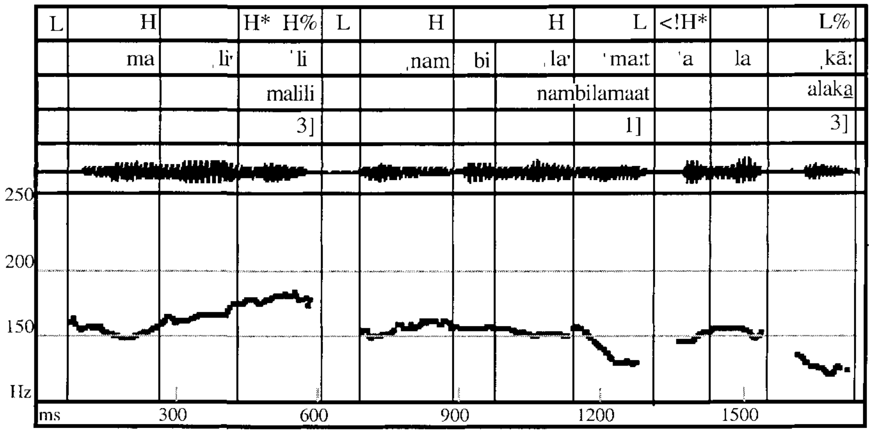


FIGURE 11.3 Utterance containing a main clause followed by a non-main clause: 'Malili nambilamaat alaka', 'S/he runs, when there's perfume here' (female speaker).

speakers use the HL% boundary tone with regularity in statements, but most employ it only rarely. The HL% boundary tone appears to be a more consistent feature of imperatives, although this observation should be regarded with some caution, since imperatives have only been elicited from two speakers. These two speakers regularly use the HL% boundary tone in imperatives but almost never in statements.

The fall from the H to the L phase of the HL% boundary tone occurs relatively late in the final syllable. This pitch fall is often imperceptible and, in many cases, can be regarded as a by-product of non-modal phonation associated with final position. However, in other tokens, the fall commences as early as the middle of the syllable and is quite perceptible. A clear example of the HL% boundary tone is provided by the statement IP in Figure 11.4. The inventory of boundary tones and the semantic contexts in which each arises are summarized in Table 11.1.

11.3.2. *Pitch accents*

Chickasaw has two types of pitch accents. The first of these accents is the morpholexical pitch accent, indicated as H^{λ}, which falls on a single syllable in certain verb forms (see Section 11.2.3). The second type of pitch accent is the nuclear pitch accent which is consistently (even in different semantic contexts) realised as a high tone, H*, falling on a syllable within the last word of an Intonational Phrase. The location of the nuclear pitch accent is predictable

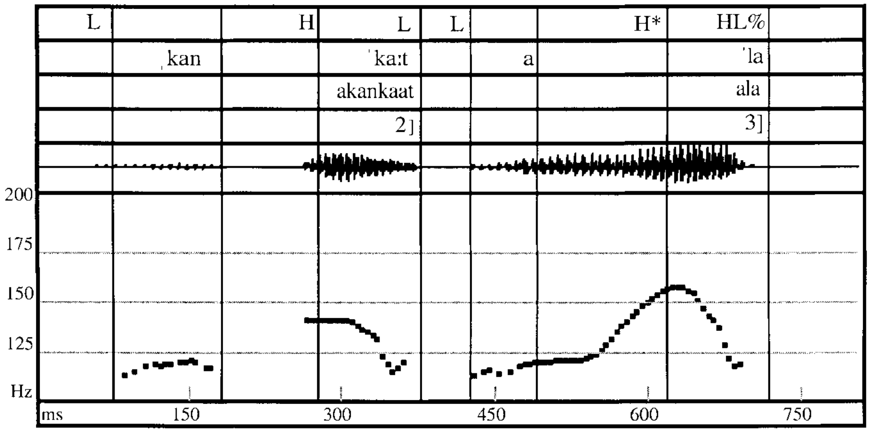


FIGURE 11.4 Final HL% boundary tone in statement IP '(A)kankaat ala', 'The skunk is here' (male speaker).

TABLE 11.1 Inventory of boundary tones

Boundary tone	Semantic context
H%, Ø%	• Statements, echo questions
HL%	• Imperatives (occasionally statements)
L%	• Wh- and yes/no questions • Exclamations • Non-main clauses, postposed nouns

and generally (with some exceptions to be discussed below) falls on either a final syllable or a heavy syllable, syllables which carry some degree of stress at the word level (see Section 11.2.2). The conditions governing nuclear pitch accent placement, however, are complex and depend on an intricate balance of phonological and morphological factors.

In statements, the final syllable of the Intonational Phrase carries the H*. In statements, the phonetic evidence for the nuclear pitch accent is less robust due to the high boundary tone often found at the end of statements (Section 11.3.1). However, in support of the nuclear pitch accent on the final syllable, the final syllable also characteristically carries not only the highest pitch but also the greatest intensity in a statement, although long vowels (including rhythmically lengthened vowels) may rival the nuclear pitch accent for having the greatest amplitude.

The H* nuclear pitch accent is more transparent in questions than in statements, due to the low boundary tone found at the end of questions

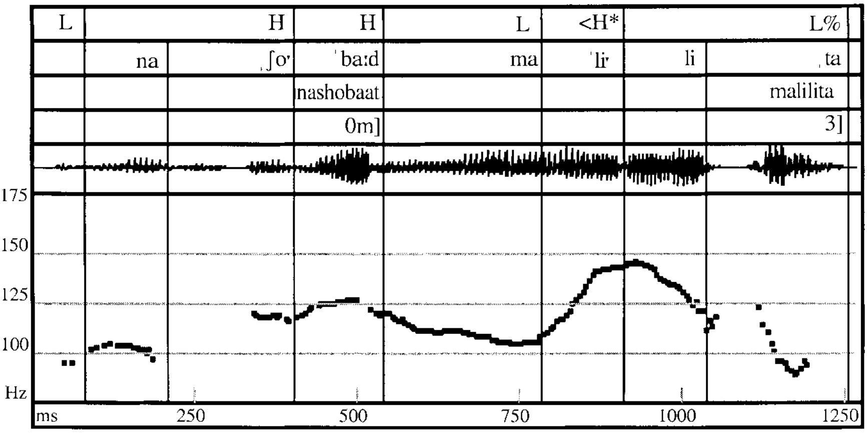


FIGURE 11.5 Nuclear pitch accent in question IP 'Nashobaat malilita', 'Does the wolf run?' (male speaker).

(Section 11.3.1). The syllable associated with the pitch peak serving as the origination point for the fall to the final low boundary tone carries the nuclear pitch accent. Often the pitch peak itself falls late in the syllable carrying the pitch accent, sometimes even during the onset of the following syllable, as in Figure 11.5, in which the pitch accent phonologically falls on the antepenult. The syllable carrying the nuclear pitch accent also has the greatest intensity in the IP.

(i) *Phonological conditions governing pitch accents in questions*: with respect to pitch accent placement, two types of questions must be distinguished. The first type of question, which is less complex in its pitch accent patterns, involves the suffixing of the interrogative marker –tõː to nouns under focus (Munro and Willmond 1994), e.g. [akankáʔtõː] 'Is it a *chicken*?'. In questions formed with –tõː, the pitch accent falls on the syllable immediately preceding the suffix and is followed by a steep fall in fo to the low boundary tone associated with the suffix. If the syllable before the suffix is open and contains a short vowel, an /h/ is typically inserted after the root-final vowel before the suffix: /falatõː/ → [faláhtõː] 'Is it a *crow*?'. The addition of the /h/ has the effect of ensuring that the final (and pitch accented) syllable of the root is heavy (see Munro 1996: 7–8 for further discussion of final /h/ in Chickasaw and closely related Choctaw).

Basic wh- and yes/no questions not formed with the noun suffix –tõː are sensitive to a more complex set of conditions governing pitch accent placement. It is these complications which we address now. We begin with the purely phonological factors and then turn to morphological factors.

A useful generalization for characterizing the location of the nuclear pitch accent in standard wh- and yes/no questions is that the transition from high pitch accent to low boundary tone minimally requires two vocalic moras, i.e. either a long vowel or two short vowels. A result of this restriction is that the only final syllable which can carry the nuclear pitch accent in a question is one containing a long vowel (CVV). Examples of final CVV carrying the nuclear pitch accent (indicated by an acute accent) appear in (4).

(4) (*a*) (katiˑmihtã) sahaˑʃáː 'Why am I angry?'
(*b*) (nantaːt) oktáːk 'What is a prairie?'
(*c*) (kataːt) maliˑtːóːk 'Who ran (distant past)?'

Phonetically, a pitch accented CVV final syllable in a question IP is characterized by a pitch peak followed by a steep pitch fall to the end of the IP. The timing of the pitch peak is crucial in distinguishing a tautosyllabic sequence of H*L% from a HL% boundary tone which can arise in statements or imperatives (see Section 11.3.1). The H* in a H*L% sequence is realized early in the final vowel, whereas the H in a HL% boundary tone is realized in the middle or towards the end of the final vowel. The early realization of the H* in H*L% is illustrated in Figure 11.6. A further difference between H*L% and HL% lies in their distributions: H*L% may not occur on a syllable containing a short vowel, whereas HL% may, as seen earlier in Figure 11.4. In addition, a L% boundary tone following H* is realized at a lower fundamental frequency than the L forming part of a HL% boundary tone. Thus, the end of the HL% boundary tone in the statement in Figure 11.4 is not

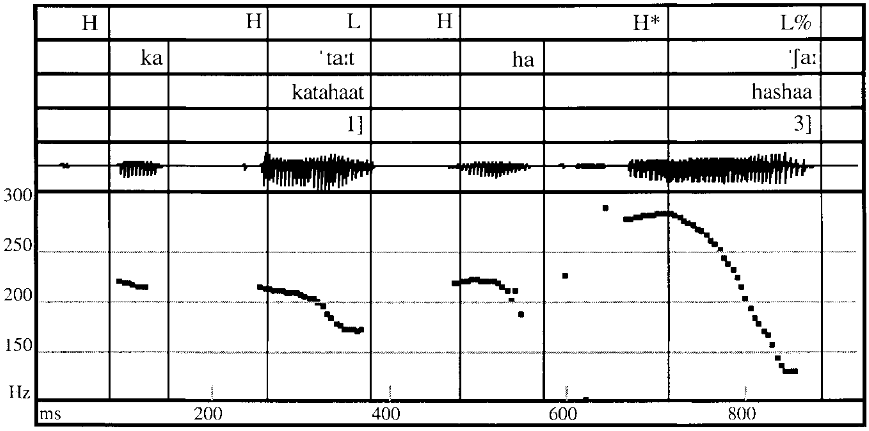


FIGURE 11.6 Final H*L% sequence in 'Katahaat hashaa?' 'Who is angry?' (female speaker).

associated with the lowest f0 of the utterance, unlike the L% boundary tone in the questions in Figure 11.6.

If the final syllable does not contain a long vowel, the nuclear pitch accent falls on the penultimate syllable if it is heavy, i.e. CVV or CVC (5).

(5) (*a*) malíːtam 'Did s/he run?'
(*b*) (nantaːt) hatáːt͡ʃim 'What turned colour?'
(*c*) (kataːt) malíˑli 'Who is running?'
(*d*) (nantaːt) tʃilákbi 'What is dry and cracked?'
(*e*) (kataːt) tokfóhli 'Who's mouth is watering?'
(*f*) (nantaːt) istóktʃank 'What's a watermelon?'

If the final syllable is not CVV and the penultimate syllable is neither CVV nor CVC, the nuclear pitch accent falls on the antepenultimate syllable (6). Because the rhythmic lengthening process (Section 11.2.1) ensures that there are not more than two consecutive syllables which are neither CVV or CVC in questions, an antepenultimate syllable carrying the nuclear pitch accent in questions will either be closed or contain a long vowel (with one morphological exception discussed in the next paragraph). Thus, any syllable carrying the nuclear pitch accent in a question IP (subject to certain exceptions discussed below) is either CVV or CVC.

(6) (*a*) mál:itam 'Did s/he jump?'
(*b*) ʃíːpata 'Is it stretchy?'
(*c*) (nantaːt) abóːkoʃiʔ 'What's a river?'

An interesting feature of the nuclear pitch accent in questions is that, unless the final syllable is CVV, it falls on a syllable which does not carry primary stress at the word-level. Thus in examples (5*d*–*f*), the pitch accent falls on a non-final syllable, even though the final syllable carries primary stress in words in phrase-medial position which lack a non-final long vowel (see Section 11.2.2). In (5*d*–*f*), the nuclear pitch accent falls on a syllable carrying secondary stress at the word-level. There are also cases in which the nuclear pitch accent falls on a syllable predicted to be unstressed at the word level. This scenario arises in phrase-final disyllabic words of the form CVCV, where the first syllable carries the nuclear pitch accent even though it is unstressed at the word-level (see Section 11.3.2). Because the nuclear pitch accented syllable is the phonetically most prominent syllable in the word in which it occurs, the pitch accented syllable is marked with primary stress in the figures throughout this paper. A relevant example of this convention is found in Figure 11.5, where the antepenult of the phrase-final words carries the nuclear pitch accent and is thus marked with primary stress.

(ii) *Morphological conditions governing pitch accents in questions*: there are also morphological conditions which interact with the purely phonological conditions governing nuclear pitch accent placement. One morphological restriction is that the nuclear pitch accent does not fall to the left of the first syllable of the root. In other words, the nuclear pitch accent is restricted from falling on prefixes, even if syllable weight conditions predict that a prefix should carry the nuclear pitch accent. For example, the verb [iliˑ-písa] in the sentence [kataːt iliˑ-písa] 'Who looks at herself/himself?' consists of the reflexive prefix iliˑ- plus the root pisa. In this form, the nuclear pitch accent falls on the penultimate syllable, /pi/, the first of the root, even though purely phonological conditions would predict that the heavy antepenultimate syllable rather than the light penult should take the nuclear pitch accent.

The nuclear pitch accent also is limited to the final word in a compound, even if this restriction means that the pitch accent falls on a syllable not predicted to carry the pitch accent on phonological grounds. For example, in the question [nantaːt akank-óʃiʔ] 'What is a chicken egg?/What is a chick?', which consists of the nouns [akankaʔ] 'chicken' and [oʃiʔ] 'baby, little one', the nuclear pitch accent falls on a syllable in the second noun of the compound, even though this syllable is light and the preceding syllable is heavy.

The nuclear pitch accent is not restricted from falling on suffixes, as the accent pattern in the question [kataːt haʃaː-tːóːk] 'Who was angry (distant past)?' indicates. In this form, the nuclear pitch accent falls on the remote past suffix -tːoːk. In fact, it is a requirement in suffixed verbs that the nuclear pitch accent fall on a syllable in the suffixal complex. Thus, in the question /pisa-li-tam/ 'Was I looking at her/him', which consists of the root [pisa] plus the first person singular suffix -li plus the question marker -tam, the nuclear pitch accent falls on the first person suffix -li rather than the syllable /sa/ which would be expected to take the pitch accent if strictly phonological conditions were observed, given that -li is a light syllable. In cases where a CV suffix carries a nuclear pitch accent against purely phonological predictions, the vowel is phonetically lengthened; thus /pisa-li-tam/ is realized as [pisaˑ-líˑ-tam]. This lengthening of the accented vowel has the effect of ensuring satisfaction of the phonological requirement that a nuclear pitch accented penult be phonologically heavy. The lengthening of the pitch accented vowel in a suffixal CV syllable contrasts with the failure of lengthening to effect vowels in root-internal CV syllables carrying a pitch accent, e.g. [kataːt tʃĩː-hójo] 'Who's looking for it for you?, in which the accented vowel in the root [hojo] remains short.

An exception to the requirement that a verbal suffix be pitch accented is provided by word-final suffixes ending in a syllable containing a short vowel: a word-final syllable must contain a long vowel to be accented. For example, in the question [katahtãː pisáˑ-li] ‘Who am I looking at?’, the first person singular suffix -li fails to attract the pitch accent because it is the final syllable and contains a short vowel. In addition to the phonological blocking of pitch accents on final non-CVV syllables, the two exclamatory suffixes, -kãː and –hV, where V is a nasalized (and therefore long) copy of the preceding vowel, reject the nuclear pitch accent in questions even though they contain a long vowel. Since the exclamation suffixes only occur as the final syllable of a word, their rejection of the pitch accent does not contradict the earlier generalization that *non-final* suffixes attract the nuclear pitch accent. It does, however, mean that we cannot know whether the absence of a pitch accent on the question markers -tam and –ta (see examples (5*a*) and (6*a*) above), which also always occur word-finally, is due to a lexically marked restriction against accented question markers or to the more general restriction against pitch accents on final short vowels. This restriction must be maintained independently as a generalization, however, since it is also observed in words lacking an overt suffix, as in (5*c*–*f*).

There is one circumstance under which two nuclear pitch accents may surface in a single IP. In an IP containing two verbs, each verb contains a syllable with a pitch accent. An example of this phenomenon is provided by the last two words in the question IP [kataːt iskánːoʔst iʃtájːa] ‘Who’s beginning to be short?’ in Figure 11.7. In this example, the first verb [iskánːoʔst] ‘be short’ carries a pitch accent on the penult, as does the second verb [iʃtájːa] ‘begin’. Typically, in cases in which two pitch accents

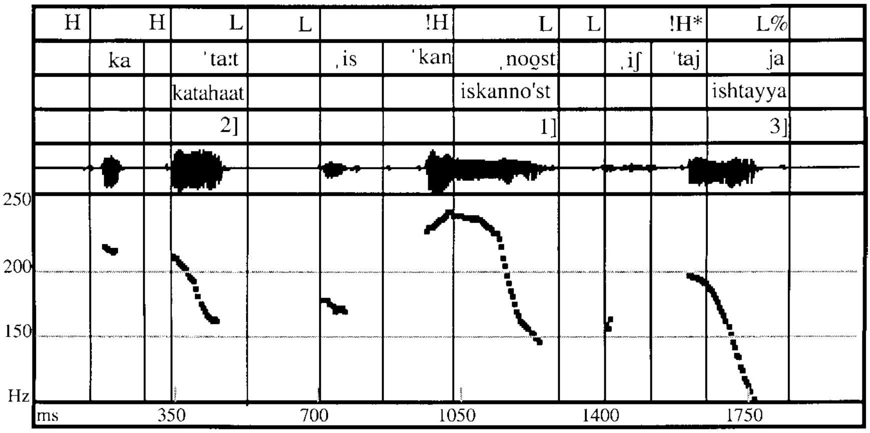


FIGURE 11.7 Two pitch accents in question IP ‘Katahaat iskanno’st ishtayya’ ‘Who’s beginning to be short?’ (female speaker).

occur in a single IP, the first pitch accent is realized with a phonetically higher pitch than the second one, as in Figure 11.7, though it is also possible for the second one to be higher. This latter scenario seems to arise frequently when the first pitch accented syllable falls on a short vowel followed by a coda obstruent.

(iii) *Interaction between morpholexical pitch accents and nuclear pitch accents*: recall from Section 11.2.3 that certain Chickasaw words carry a morpholexical pitch accent on one syllable. These morpholexical pitch accents are phonetically realized as high tones parallel to the phonologically predictable nuclear pitch accents.

A single word may have both a morpholexically pitch accented syllable and another syllable with a nuclear pitch accent. In practice, however, this situation occurs rarely, since there is a restriction against a high tone (whether due to nuclear pitch accents, morpholexical pitch accents, or boundary tones) on a syllable adjacent to a morpholexically accented syllable. In case phonological (or morphological) conditions would predict that a H* nuclear pitch accent or a H% (or HL%) boundary tone would fall on a syllable immediately adjacent to the morpholexically pitch accented syllable, this other high tone is not realized. This restriction only pertains to high tones adjacent to morpholexically pitch accented syllables, since, as we have already seen, tautosyllabic sequences of H* and H% are permissible in statements.

Figure 11.8 contains an example of a statement, [tʃofânta] 'S/he is cleaner' n-grade, in which the only high tone is the morpholexical pitch accent on the penultimate syllable. The H* pitch accent normally found on final syllables in statements is suppressed and the terminus of the statement is marked by a relatively flat mid-level fundamental frequency plateau through the final syllable. To differentiate morpholexical pitch accents from phonological ones, morpholexical ones are transcribed in figures with a superscripted Greek letter lambda following the high tone.

The phonological analysis of the final pitch plateau following a morpholexically accented penult is somewhat problematic, since it differs substantially from the realization of the three boundary tones posited thus far: L%, H%, and HL%. In fact, statements and questions ending in a verb containing a morpholexical pitch accent on the penult are differentiated solely on the basis of the final pitch excursion. Statements have a flat mid level plateau, whereas questions have a fall.

One possible analysis of the level plateau found in statements would assume that the plateau in Figure 11.8 reflects the absence of a boundary tone, i.e. Ø%, where a boundary lacking a phonological tone is phonetically

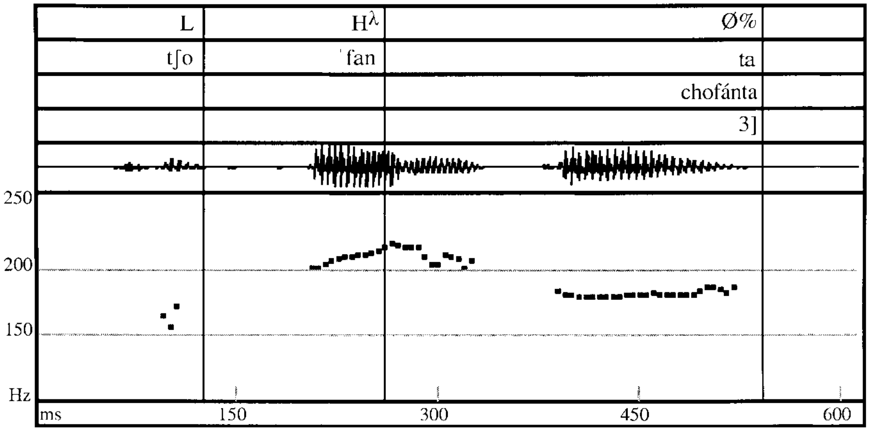


Figure 11.8 Morpholexical pitch accent in IP 'Chofánta', 'S/he is cleaner' (female speaker).

realized as a mid tone. Another possibility is to analyse the final mid tone as the phonetic manifestation of a downstepped H%, i.e. !H%, though one might expect !H% to be realized with slightly higher fundamental frequency than is typical in examples like the one in Figure 11.8. I will tentatively adopt the Ø% boundary tone transcription here with the caveat that future research may argue for an alternative analysis.

The nuclear pitch accent also does not fall on a syllable adjacent to a morpholexically accented syllable in questions. For example, in the question IP [tʃofâjːaʔtata] 'Is she/he really clean?' y-grade, the presence of the morpholexical pitch accent on the preantepenult precludes a nuclear pitch accent on the antepenult, as illustrated in Figure 11.9.

If, however, there is at least one syllable separating the morpholexically pitch accented syllable from the potential docking site for the nuclear pitch accent and/or the final boundary tone, then both the morpholexical pitch accent and the nuclear pitch accent and/or boundary tone are realized. For example, in the statement IP consisting of the word [hîkːiʔjá] 'S/he is standing', the morpholexical pitch accent falls on the initial syllable and the nuclear pitch accent and boundary tone falls on the final syllable. Figure 11.10 illustrates the question [hôjːoʔlolíˑta] 'Am I wearing shoes?', in which the first syllable carries the morpholexical pitch accent while the penultimate syllable attracts the nuclear pitch accent. In this particular example, the nuclear pitch accent falls on a suffix, -li 1sg subject. The nuclear pitch accent is realized with

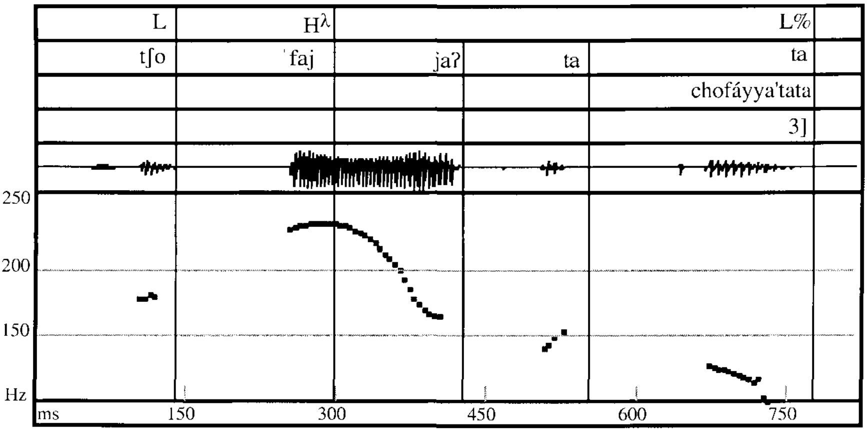


FIGURE 11.9 Morpholexical pitch accent in 'Chofáyya'tata', 'Is s/he really clean?' (female speaker).

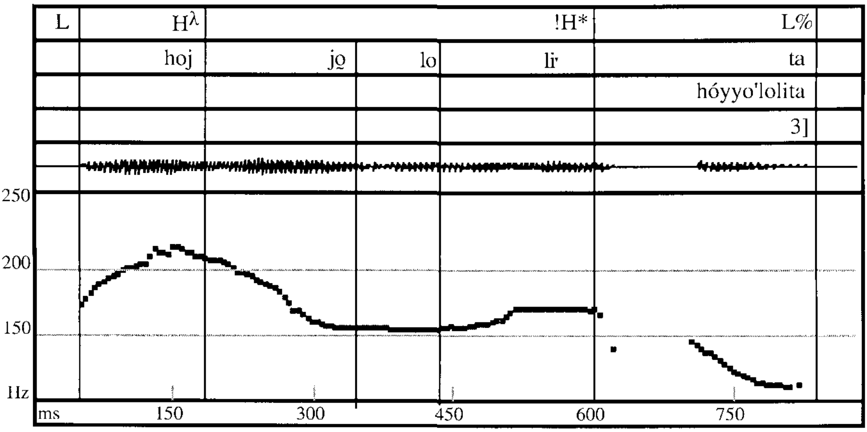


FIGURE 11.10 Morpholexical and nuclear pitch accents in 'Hóyyo'lolita', 'Am I wearing shoes?' (female speaker).

a much lower fundamental than the morpholexical pitch accent; this lowering of the nuclear pitch accent is reflected in the transcription of the nuclear pitch accent as a downstepped !H* (see Section 11.5.1 for discussion of downstep). It is also worth noting that there is characteristically a sag in pitch between a morpholexical pitch accent and a following nuclear pitch accent in the same word.

11.4. PROSODIC STRUCTURE

Evidence suggests the existence of at least three constituents at or above the level of the prosodic word in Chickasaw: the Intonational Phrase (Section 11.4.1), the Accentual Phrase (Section 11.4.2), and the Prosodic Word (Section 11.4.3). In addition, there is a domain smaller than the Prosodic Word in which the phonological process of rhythmic lengthening applies (Section 11.4.4).

11.4.1. *Intonational phrase*

In Chickasaw, the largest prosodic constituent which is defined intonationally is the Intonational Phrase (IP). Most sentences consist of a single IP which contains one or more nuclear pitch accented syllables (see Section 11.3.2) and a boundary tone at its right edge (see Section 11.3.1). The right edge of the Intonational Phrase is also often associated with non-modal phonation, either breathiness or creakiness. Creakiness is particularly common as a by-product of the L% boundary tone in questions (see Figures 11.2, 11.5, 11.6, 11.9).

An exception to the generalization that a sentence consists of a single IP is provided by sentences in which the canonical Chickasaw SOV word order is substituted with an order in which a noun rather than a verb appears in sentence final position. In such sentences, the verb ends the first IP and any postposed noun phrases form a separate IP (see Section 11.5.1). A similar splitting of an utterance into two IPs is found in biclausal utterances, in which each clause characteristically constitutes its own IP (see Section 11.3.2).

11.4.2. *Accentual phrase*

A single IP is in turn composed of one or more intonational units, which may be termed Accentual Phrases (AP). Each AP is defined tonally by a series of tones aligned with different prosodic positions within the AP. The degree of perceived disjuncture at an AP boundary is smaller than at an IP boundary.

An AP may consist of more than one morphological word, where a morphological word refers to a root plus all bound affixes. Conversely, a long morphological word may consist of more than one AP. Thus, in Chickasaw the Accentual Phrase can be either smaller or larger than the morphological word (see Section 11.4.5 for further discussion). There is a strong preference for AP boundaries to coincide with morphological word boundaries. Thus, each morphological word characteristically is a single AP and each AP typically consists of a single morphological word. A by-product of this strong tendency toward alignment of AP and morphological word boundaries is that a sequence of two CVCV words is characteristically treated as two separate APs rather than one. The likelihood of two morphological words being produced as a single AP is greater for words which are constituents. For example, a sequence of object followed by a verb is more likely to be realized as a single AP than a sequence of subject plus verb or subject plus object. An example of two morphological words forming a single IP appears in Figure 11.11, in which the subject [minkaːt] and the verb [ala] coalesce to form a single AP. The final stop of the noun is flapped in this example, a phenomenon affecting word-final alveolar stops in AP-medial position at fast speech rates (see Section 11.4.2 for segmental diagnostics for AP-phrasing).

In general, the likelihood of any sequence of two morphological words being uttered as a single AP is small. In cases of multiple morphological words forming a single AP, there are typically segmental correlates associated with the intonational parse, as in Figure 11.11 (see Section 11.4.2).

On the other hand, long morphological words (those greater than seven syllables) may optionally consist of more than one AP, where the likelihood of the morphological word being divided into multiple APs increases commensurately with the length of the word. For example, the nine syllable word /akitːimanompoʔlokitok/ 'I didn't speak to him' may be divided into two

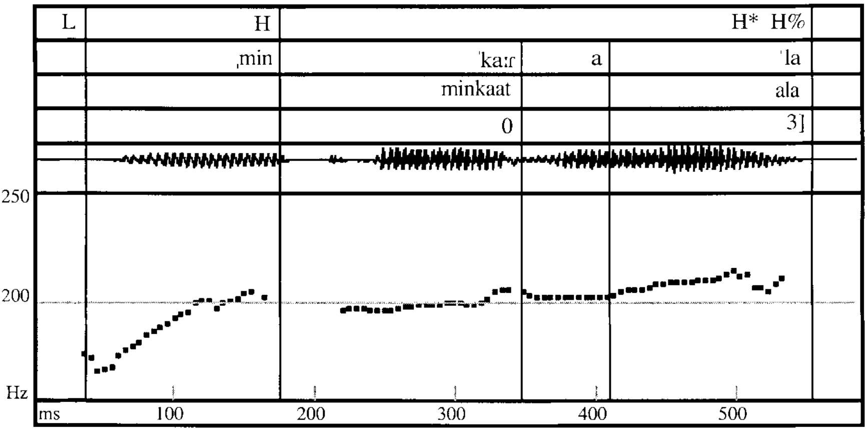


Figure 11.11 Two morphological words forming a single Accentual Phrase in 'Minkaat ala', 'The chief is here' (female speaker).

APs consisting of a six syllable AP followed by a three syllable AP, i.e. [akitːimaˈnompoʔ]AP[lokiˈtok]AP. The division of longer morphological words into multiple APs appears to be sensitive to the morphological composition of a word, though the details of this influence of morphology on the intonational parse require further investigation.

(i) *Tonal realization of the Accentual Phrase*: in describing the tonal realization of the AP, it is useful to invoke the notion of the mora, where, in Chickasaw, a short vowel or a sonorant coda consonant are each associated with one mora and a long vowel is associated with two moras. Onset consonants and coda obstruents are non-moraic for purposes of Accentual Phrase tonal alignment in Chickasaw (though coda obstruents contribute to the weight of a syllable for purposes of pitch accent placement in non-final position). The canonical realization of the AP pattern is [LHHL], a pattern which is typical of APs containing at least three moras. The [LHHL] pattern is also a marked option in shorter APs, where the likelihood of all tones being realized decreases as the duration of the AP shortens. The realization of the AP tones in a short AP is discussed below. Syllables carrying H AP tones are not reliably associated with greater intensity and duration than other syllables; this separation of intensity and duration from fundamental frequency is thus diagnostic in distinguishing between AP tones and both IP level morpholexical and phonological pitch accents.

In APs in which all four tones are realized, the initial low is associated with the left edge of the AP. The first high tone occurs fairly early in the AP; it is generally realized on the second mora. Thus, if the first syllable of a word contains a long vowel or is closed by a sonorant, i.e. if the first syllable is bimoraic, the first high tone is usually realized on the first syllable. If, however, the first syllable contains only one mora, the high is delayed until the first mora of the second syllable. The actual timing of the first high tone is only loosely linked to the number of sonorant moras. If, for example, a long vowel in the first syllable is phonetically shortened, as at a faster speech rate, the high tone may actually fall on the second syllable rather than the first one, as in Figure 11.4.

The second high tone is loosely associated with the beginning of the final syllable of the AP. Syllables intervening between the two high tones are realized with high tone by interpolation. The final high tone is usually followed by a sharp fall in pitch to the final low tone associated with the right edge of the AP. The realization of the AP tones is shown in schematic form in (7). An AP with a canonical tonal realization, including a lengthy high-toned plateau, is illustrated in Figure 11.12.

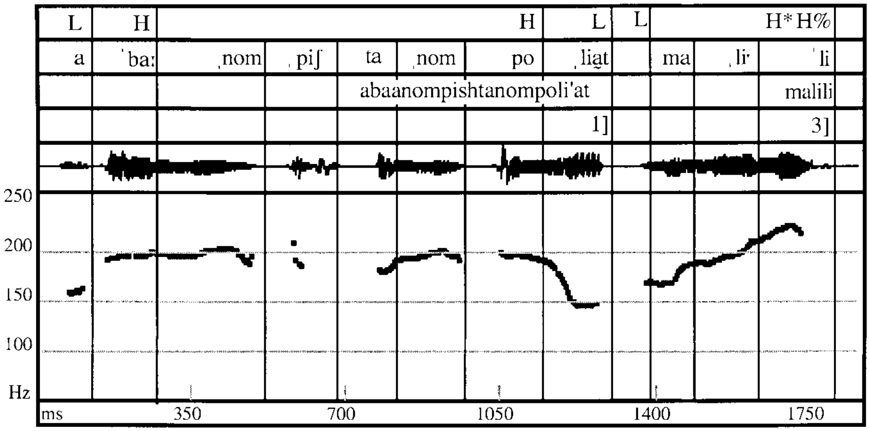
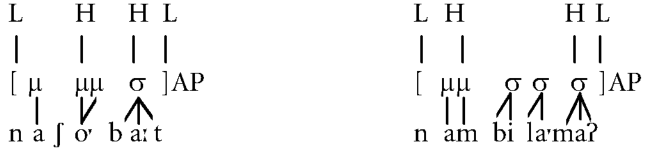


FIGURE 11.12 Accentual phrase tones in 'Abaanompishtanompoli'at malili', 'The preacher runs' (female speaker).

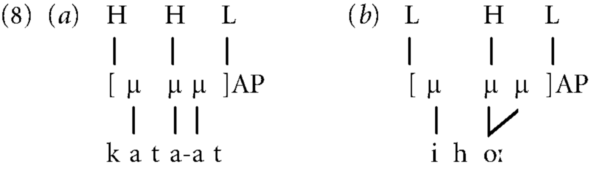
(7) (*a*) Monomoraic first syllable (*b*) Bimoraic first syllable



The final pitch fall is not an invariant property of the AP. If the final syllable contains a long vowel, which carries primary word-level stress (see Section 11.2.2), the pitch fall on the final syllable may optionally be absent and replaced with a high level plateau, which may be interrupted by a local fo peak. The boost in fo and the suppression of the AP final L tone often triggered by word-final long vowels is evident in Figure 11.5 on the primary stressed syllable /ba:d/ in the first word [na,ʃo·'ba:d].

The canonical AP tonal pattern is often truncated in an AP which contains fewer than three moras. The most common tonal pattern in a short AP is [HL] with the H realized at the beginning of the AP and the L at the right edge of the AP. The result is a steady fall in pitch throughout the AP. This [HL] intonation pattern is consistent with the observation that many disyllabic words may be realized with prominence on the first rather than the final syllable (Munro 1996). The tonal realization of the AP is sensitive to the morphological structure of words, albeit in different ways than IP-level pitch accents. For purposes of determining the ability of a word to manifest the full tonal realization of the AP, suffixes are ignored. This contrasts with IP-level pitch accents which are attracted by suffixes (Section 11.3.2). The null contribution of suffixes to the mora count of the AP is illustrated in Figures 11.6 and 11.7 in which the word [kata-at] 'who' consists of a bimoraic root kata- plus a monomoraic suffix -at. In these examples, the AP [kataːt] fails to realize the initial L even though the AP contains a total of three moras, as shown schematically in (8*a*). This contrasts with a monomorphemic trimoraic AP, which realizes the initial L, as in the AP [ihoː] 'woman' (8*b*).

(8)



The initial L may also be suppressed or realized at a slightly higher level if there is H* or a morpholexical pitch accent in the immediate vicinity, either on the initial syllable or early in the peninitial syllable. In Figure 11.6, the initial low tone of the second AP consisting of the word [haʃaː] does not surface due to the realization of the H* at the beginning of the second vowel. This contrasts with Figure 11.7 in which the initial L of the AP [iʃtajːa] is realized despite the H* realized early in the peninitial syllable.

Another more marked option in a short AP is to not realize either of the AP high tones; the result is a level low toned AP with a L linked to the beginning of the AP and a L associated with the end of the AP. The various realizations of the AP are summarized in Table 11.2.

(ii) *Segmental diagnostics for the Accentual Phrase*: certain segmental phenomena applying in rapid speech are typically bounded by the Accentual Phrase and thus can be employed as diagnostics for determining whether Accentual Phrase boundaries are present or not. Some of these segmental diagnostics include the following. An intervocalic alveolar stop before a stressless vowel optionally undergoes flapping at fast speech rates in AP medial position, including across word boundaries, as in Figure 11.11 (9*a*). An intervocalic consonant may resyllabify with a following vowel across a word-boundary, with accompanying aspiration in the case of stops (9*b*). An

Table 11.2 Tonal realizations of the Accentual Phrase

Tonal pattern	Condition
LHHL	AP $\geq$ 3 moras (excluding suffix)
HL	AP $<$ 3 moras
LL	AP $<$ 3 moras (less common)
LHH	Last σ is CVV (optional)

intervocalic stop consonant may undergo voicing between sonorants (9*c*, *d*). In addition, vowels, particularly high vowels, optionally syncopate in hiatus contexts across word boundaries in AP medial position (9*e*).

(9) (*a*) ˌminˈkaː**t** aˈla → ˌminˈkaː**ɾ** aˈla ‘The chief is here.’

(*b*) ˌhat.ˈːa**k**. a.ˌpiˑ.ˈla → ˌhat.ˈːa.**k**^{h} a.ˌpiˑ.ˈla ‘S/he helps the man.’

(*c*) ˌminˈkaː**t** aˈla → ˌminˈkaː**d** aˈla ‘The chief is here.’

(*d*) tʃiˌpoˈtaː**t** ˈjaː → tʃiˌpoˈtaː**d** ˈjaː ‘The child is crying.’

(*e*) faˈla ˌ**i**ʃˈkin → faˈla ʃˈkin ‘the crow’s eye’

It should be noted that these diagnostics, though relatively reliable in diagnosing whether an AP boundary is present or not, are not foolproof. Thus, there are instances of the processes exemplified in (9) occurring across an AP boundary (see Section 11.5.4 for transcription guidelines for these mismatches between segmental diagnostics and prosodic constituency).

11.4.3. *Prosodic word*

The Prosodic Word is coextensive with the morphological word and is the domain of word-level stress assignment. The need to separate the Prosodic Word, the domain of stress assignment, from the rhythmic lengthening domain (Sections 11.2.1, 11.4.4), a domain smaller than the Prosodic Word, becomes apparent under two circumstances. First, prefixes which fall outside of the domain of rhythmic lengthening are eligible to carry primary stress if they possess the requisite phonological properties, i.e. if they contain the rightmost long (or lengthened) vowel in a word. Second, in compounds in which rhythmic lengthening is blocked across the boundary between the two words forming the compound, the primary stress for the compound as a whole still falls on a syllable in the first word of the compound provided it contains the rightmost long (or lengthened) vowel in the compound.

11.4.4. *Rhythmic lengthening domain*

Like nuclear pitch accent placement, rhythmic lengthening is also sensitive to morphology. The domain of rhythmic lengthening is discussed in considerable detail in Munro and Willmond (1994) and Munro (1996) and is smaller than the domain defined in Section 11.4.3 as the Prosodic Word. Certain prefixes fall outside the domain of rhythmic lengthening and are disregarded in the syllable count used to determine which vowels are lengthened. Thus, in the word [t͡ʃim-abiˑtok] 'S/he killed her/him for you', the prefix tʃim- falls outside the rhythmic lengthening domain, which commences with the first vowel of the root. Certain suffixes, including all noun suffixes, also fall outside of the rhythmic lengthening domain and, for still other (disyllabic) prefixes, the second but not the first syllable is part of the rhythmic lengthening domain (see Munro and Willmond 1994 for discussion of these complications and others).

11.4.5. *Summary of the hierarchical organization of prosodic constituents*

In summary, there are at least four hierarchically arranged prosodic constituents in Chickasaw: from largest to smallest, the Intonational Phrase, the Accentual Phrase, the Prosodic Word, and the Rhythmic Lengthening Domain. Examples illustrating the relationship between these constituents and the morphological word appear in (10).

(10) Examples of prosodic constituency in Chickasaw

```
(a)
Intonation Phrase   {                                  }
Accentual Phrase    {                    }{            }
Morphological Word  {                                  }
Prosodic Word       {                                  }
Rhythmic Length              {                         }
                    aː t͡ʃ i m a b iˑ k a t͡ʃ iˑ l i t o k  'I made him sick there for you.'
                    aː-   t͡ʃim- abika- t͡ʃi-   li- tok
                    there you    sick     caus  1sg past

(b)
Intonation Phrase   {                 }{            }
Accentual Phrase    {                 }{            }
Morphological Word  {         }{      }{            }
Prosodic Word       {         }{      }{            }
Rhythmic Length     {         }{      }{        }
                    naʃoˑba     pisa    tʃipoˑtaːt  'The child looks at the wolf.'
                    naʃoba      pisa    tʃipoˑta-at
                    wolf        sees    child-subj.
```

The form in (10*a*) consists of a single Morphological Word isomorphic to the Prosodic Word. There are two prefixes aː- and t͡ʃim- outside the rhythmic lengthening domain, which consists of the root plus the suffixes. The form in (10*a*) also contains two tonally defined Accentual Phrases whose boundaries coincide with those of neither the Prosodic nor the Morphological Word. Finally, the two Accentual Phrases are grouped into one Intonational Phrase.

The SVO sentence in *(lOb)* groups the first two Morphological Words, the object and the verb, each of which is coextensive with a Prosodic word and a Rhythmic Lengthening Domain, into a single Accentual Phrase, exactly the opposite of *(lOa)* in which one Morphological Word is divided into two Accentual Phrases. The postposed subject in *(lOb)* forms an Intonational Phrase independent of the object and verb, which together constitute a Intonational Phrase. The subject suffix in the postposed noun falls outside the Rhythmic Lengthening Domain comprising the root. Although the examples in (10) do not exhaust all possible constituencies in Chickasaw, they serve to illustrate some of the range of variation in Chickasaw prosodic structure.

11.5. A TRANSCRIPTION SYSTEM

The system proposed here makes use of four tiers for transcribing Chickasaw intonation. These tiers, from top to bottom, are as follows. The Tone Tier provides an intonational analysis of an utterance (Section 11.5.1). The Phonetic Transcription Tier provides an IPA transcription of the utterance (Section 11.5.2). The Orthographic Tier gives the orthographic version of the utterance (Section 11.5.3). Finally, the Break Index Tier provides numerals indicating the relative level of disjuncture between constituents (Section 11.5.4). Labels used within each of the tiers are summarized in the Appendix.

11.5.1. *The tone tier*

The tone tier consists of the intonational analysis of an utterance, including the pitch accents, both phonological and morpholexical, the Accentual Phrase tones, and the Intonational Phrase boundary tones. The tonal transcription may be supplemented with diacritics in certain cases. There are two circumstances discovered thus far in which diacritics are useful in providing a detailed transcription of the phonetic realization of the tones. First, placing the diacritic < before a pitch accent indicates that the pitch peak associated with the accent is realized after the syllable which is phonologically associated with the accent, as in Figure 11.5.

Another important diacritic is the downstep diacritic associated with downstepped pitch accents. There are a number of circumstances in which downstepped pitch accents are likely to be transcribed. First, the downstep diacritic ! may be used before a phonological pitch accent which is phonetically lower than a preceding pitch accent in the same Intonational Phrase, as in

Figure 11.7. It is also possible to use the downstep diacritic when a phonological pitch accent is preceded by a phonetically higher morpholexical pitch accent in the same IP, as in Figure 11.10. A final context in which a downstepped pitch accent commonly occurs is in an **IP** associated with tonal lowering and compression, as in postverbal nouns and postposed non-main clauses, as in Figure 11.3.

11.5.2. *The phonetic transcription tier*

The phonetic transcription tier contains information about the segmental properties of an utterance, as well as prosodic features other than intonation, including stress. Because the transcription is phonetic rather than strictly phonological, it includes information about stress as well as allophonic segmental information, such as the processes discussed in Section 4.2.2.

11.5.3. *The orthographic tier*

The orthographic tier provides a transcription of the Chickasaw orthographic system, which encodes phonemic contrasts at the segmental level, including length, and also morpholexical pitch accents (typically indicated by an acute accent). The letter-to-sound correspondences in Chickasaw closely resemble their counterparts in English with certain exceptions: the symbol 'lh' represents a voiceless lateral fricative, an apostrophe indicates a glottal stop, and an underlined vowel stands for a nasalized vowel. Long vowels are written double. The interested reader is referred to Munro and Willmond (1994: ix-xi) for discussion of the orthography.

11.5.4. *The break index tier*

The break index tier provides information about the level of disjuncture between adjacent elements. The perception of disjuncture may be cued by several properties, such as pauses, non-modal phonation at prosodic boundaries, segmental effects such as final lengthening or initial fortition, among other acoustic features. At this stage in the investigation of Chickasaw, it is useful to invoke four break indices expressed as numerals ranging from '0' to '3' where a higher number is associated with a greater degree of perceived disjuncture. Thus, the boundary of a large intonational constituent such as an **IP** is marked with a higher break index value than the boundary of a smaller intonational constituent such as an AP or Prosodic Word.

Because of the large number of Accentual Phrase tones transcribed in most Chickasaw utterances, boundaries between Accentual Phrases are indicated by brackets in the Break Index Tier in order to facilitate interpretation of the membership of individual tones. If an AP boundary coincides temporally with a numerical break index, this is indicated by a bracket immediately following the break index.

(i) *The break indices:* the lowest break index is 0 which indicates a very close juncture between two prosodic elements. Prosodic constituents separated by 0 are perceived to be prosodically a single unit though intonational evidence may indicate that the two constituents are distinct. A break index of 0 may be used to mark the boundary between two Accentual Phrases forming a single Morphological Word. In the latter case, a diacritic would be used to indicate the mismatch between the close segmental juncture between the two Accentual Phrases and the relatively large intonationally defined AP break (see Section 11.5.4 (ii) for diacritics).

A break index of 1 indicates a greater degree of disjuncture than 0. The value 1 is usually associated with at least a small pause between prosodic elements. A break index of 2 indicates a greater degree of disjuncture than 1, often cued by a small amount of final lengthening and a longer pause between constituents. Break indices of 1 and 2 are typically associated with the boundary between two APs belonging to different morphological words. The difference between a break index of 1 and 2 is rather subtle, but seems warranted given the difference in perceived level of disjuncture between a subject and the object in an SOY sentence compared to the disjuncture between an object and the verb in a SOY sentence. The perceived level of disjuncture between the subject and object is characteristically greater than the degree of disjuncture perceived between an object and verb. In the proposed system, this difference is transcribed primarily as a difference in break indices: 1 for the object-verb boundary and 2 for the subject-object boundary. It is conceivable that this difference in disjuncture reflects an additional level of prosodic constituency which would group the object and verb together to the exclusion of the subject, following the expected syntactic constituency. Determination of the exact prosodic properties defining this hypothesized constituent must await further research.

The final break index is 3, which is almost always associated with an IP boundary. A number of prosodic properties signal a break index of 3. The fundamental frequency range is reset, there is a lengthy pause following the boundary, and the segment before the boundary is lengthened to some extent and may be realized with non-modal phonation.

(ii) *Diacritics in the break index tier:* in general, there is a close match between intonationally defined constituents and break indices. The value of 0 is usually used for constituents belonging to the same AP, the values of 1 and 2 are typically associated with the boundary between two APs, and the value of 3 characteristically corresponds to an IP boundary. Given this link between break indices and intonationally defined constituents, it is useful to have a notation to indicate exceptional cases in which the break index typically found at an intonationally defined boundary is substituted with another index. Such mismatches between intonational constituent and break indices might arise under several circumstances. For example, an IP boundary might be associated with a break index of 2 rather than 3. Similarly, an AP boundary might be associated with a break index of 3 in case of an unusually long pause after the AP, or 0 in case of a AP break within a Morphological Word. In such mismatch cases, a lower case 'm' can be used after the break index value to signal a mismatch between intonational constituency and constituency defined in terms of more general perceived degree of disjuncture, including segmental diagnostics. An example of the mismatch diacritic appears in Figure 11.5, in which there is a mismatch between the AP boundary, usually associated with a 1 or 2 break index value but here marked with a 0 value, and the perceived close degree of juncture between the words [naJo'bard] and [mali'lita], attributed in large part to the application of intersonorant voicing to the final alveolar stop in the first word.

11.6. SUMMARY AND CONCLUSIONS

In summary, this chapter has provided a description of and means for transcribing the principal features of Chickasaw intonation. The present study is clearly not an exhaustive study of Chickasaw intonation; such a study would require investigation of many other features, such as the use of intonation to convey various types of semantic and pragmatic information, quantitative aspects of the timing of fundamental frequency events, and the effects of speech rate on intonation. Investigation of these properties and others will doubtless necessitate modifications and additions to the model proposed here. It is hoped, however, that this study provides a general framework for describing the intonational system of a language whose prosodic system differs rather substantially from others whose intonational systems have been formally modelled.

APPENDIX: SUMMARY OF CHICKASAW INTONATIONAL LABELS

H*	*Nuclear pitch accent:* falls on a syllable in the rightmost word of the IP.
H^{λ}	*Morpholexical pitch accent:* lexically marked pitch accent in certain words.
!H*	*Downstepped pitch accent:* pitch accent with lowered fo peak relative to an earlier pitch accent within the same IP.
<	*Late Fo event:* marked on the actual Fo peak when it occurs after the syllable bearing the phonological pitch accent.
H%	*Boundary tone:* occurs at the end of statements and echo questions.
0%	*Boundary tone:* occurs at the end of statements.
L%	*Boundary tone:* occurs at the end of wh- and yes/no questions, non-main clauses, exclamations, and postposed nouns.
HL%	*Boundary tone:* occurs at the end of imperatives.
H,L	*Accentual Phrase tones:* aligned with different positions in the AP.
o	*Break index:* indicates strong cohesion, often used to signal an AP boundary within a morphological word.
	Break index: indicates moderate cohesion, often used to signal an AP boundary between object and verb.
2	*Break index:* indicates moderate lack of cohesion, often used to signal an AP boundary between subject and object.
3	*Break index:* indicates strong lack of cohesion, often used to signal IP boundary.
m	*Mismatch:* marked after the break index value to indicate a mismatch between tones and the degree of disjuncture.

REFERENCES

BECKMAN, M. E., and HIRSCHBERG, J. (1994), 'The ToB! Annotation Conventions', ms, Ohio State University.

——, and PIERREHUMBERT, J. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook,* 3: 255-309.

——, HIRSCHBERG, J., and SHATTUCK-HuFNAGEL, S. (this volume Ch. 2), 'The Original ToB! System and the Evolution of the ToB! Framework'.

BISHOP, J., and FLETCHER, J. (this volume Ch. 12), 'Intonation in Six Dialects of Bininj Gun-Wok'.

CAMPBELL, N., and VENDITTI, J. (1995), 'J-ToB!: An Intonational Labelling System for Japanese', paper presented at the Fall 1995 meeting of the Acoustical Society of America, St Louis, MO, 27 NOV.-1 Dec.

GODJEVAC, S. (this volume Ch. 6), 'Transcribing Serbo-Croatian Intonation'.

——, MUNRO, P., and LADEFOGED, P. (2000), 'Some Phonetic Structures of Chickasaw', *Anthropological Linguistics,* 42: 366-400.

JUN, S.-A. (1993), 'The Phonetics and Phonology of Korean Prosody', Ph.D. dissertation (Ohio State University) [published in 1996 by New York: Garland].

—— (this volume Ch. 8), 'Korean Intonation and Prosodic Transcription'.

——, and FOUGERON, C. (1995), 'The Accentual Phrase and the Prosodic Structure of French', in *Proceedings of XIIIth International Congress of Phonetic Sciences* (Stockholm, Sweden), Vol. 2: 722-5.

MUNRO, P. (1996), 'The Chickasaw Sound System', Ms (Los Angeles: UCLA).

—— (forthcoming), 'Chickasaw', in H. Hardy and J. Scancarelli (eds.), *Native Languages of the Southeastern United States* (Lincoln: University of Nebraska Press).

——, and ULRICH, C. (1984), 'Structure-preservation and Western Muskogean Rhythmic Lengthening', *West Coast Conference on Formal Linguistics,* 3: 191-202.

——, and WILLMOND, C. (1994), *Chickasaw: An Analytical Dictionary* (Norman: University of Oklahoma Press).

——, —— (1999), *Chikashshanompa' Kilanompoli'* (Los Angeles: UCLA Academic Publishing Service).

PIERREHUMBERT, J. (1980). 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology) (published, Bloomington, IN: Indiana University Linguistics Club).

——, and BECKMAN, M. (1988), *Japanese Tone Structure (Linguistic Inquiry Monograph,15)* (Cambridge, MA.: MIT Press).

PITRELLI, J., BECKMAN, M., and HIRSCHBERG, J. (1994), 'Evaluation of Prosodic Transcription Labelling Reliability in the ToB! Framework', in *Proceedings of the 1994 International Conference on Spoken Language Processing* (Yokohama, Japan), 1: 123-6.

SILVERMAN, K., BECKMAN, M., PITRELLI, J., OSTENDORF, M., WIGHTMAN, C., PRICE, P., PIERREHUMBERT, J., and HIRSCHBERG, J. (1992), 'ToB!: A Standard for Labelling English Prosody', *Proceedings of the 1992 International Conference on Spoken Language Processing,* Vol. 2: 867-70.

VENDITTI, J. (1995), 'Japanese ToB! Labeling Guidelines', ms, Ohio State University.

—— (this volume Ch. 7), 'The LToB! Model of Japanese Intonation'.

12

Intonation in Six Dialects of Bininj Gun-wok

Judith Bishop and Janet Fletcher

12.1. INTRODUCTION

The intonational systems of most non-Indo-European languages have been poorly studied relative to those of languages such as English, Swedish, German, or Dutch, for example. Yet it is particularly significant to examine the intonational systems of typologically diverse languages in light of renewed interest in 'intonational universals' (e.g. Vaissiere 1995). A handful of intonational studies of languages as varied as Bengali (Hayes and Lahiri 1991), Balinese (Hermann 1997), and Chickasaw (Gordon, this volume Ch. 11) have appeared in the last ten years. Coupled with this is the growing interest in documenting intonational variation within dialects of a language. The analyses of the Venlo dialect of Dutch (Gussenhoven and van der Vliet 1999) and research on Swedish dialects (Bruce *et ai.* 1999; this volume) are a welcome development.

It is not surprising therefore that few indigenous Australian languages have significant intonational descriptions, with the exception ofDyirbal (King 1992, 1994) a language once spoken in Northern Queensland, for which there are no remaining speakers; Warlpiri (King 1999), spoken in Central Australia; and Wik-Mungkan (Sayers 1974), spoken on the Cape York Peninsula, Queensland. In this chapter we examine the intonational phonology of six closely related varieties of a Northern Australian language, Bininj Gun-wok, also known as Mayali. We then outline transcription conventions which are

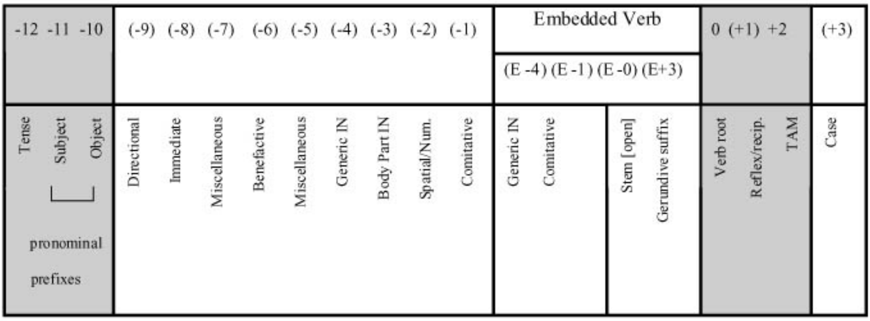
This work was carried out with the assistance of: an Australian Research Council Large Grant to the second author; Associate Professor Nicholas Evans, *Intonation, Prosody and Meaning in Australian Languages* (1998-2000); and an Australian Postgraduate Research Award and Faculty of Arts Fieldwork Grant to the first author. We wish to express our gratitude to all our language consultants, to Nicholas Evans for comments on earlier drafts of the paper, and to Murray Garde for providing us with recordings from Kuninjku.

designed to transcribe significant prosodic events in this language and its various dialects. The dialects are Gun-djeihmi, Kundedjnjenghmi, Kune, Kunwinjku, Kuninjku, and Manyallaluk Mayali.

Our study is located within the autosegmental-metrical (AM) intonational framework developed by Bruce (1977), Pierrehumbert (1980), Beckman and Pierrehumbert (1986), and Ladd (1996). According to this framework, and others such as the 'British School' of intonation (e.g. Halliday 1967), intonation performs a basic delimitative function across languages. In other words, there are tones that appear to perform some kind of phrasal edge-marking function in spoken language. Languages such as English, Japanese, Swedish, and French display various kinds of right-boundary-marking tone, i.e., a fall or rise that defines the right edge of a phrase. These phrase edges are often accompanied by other junctural phenomena, namely lengthening, pause, or sandhi-blocking (Beckman 1996). In addition, the left edge of a constituent may be marked by a sharp rise in pitch. In languages such as Balinese (Hermann 1997), Korean Oun 1993), and French Oun and Fougeron 1995; Vaissiere 1995), left and right edge boundary tones are the main indicators of intonational constituency.

According to Beckman (1996), many languages also have some kind of tonal event that is linked to a syllable or mora and performs a prominence-enhancing function within the domain of an intonational phrase. In many intonational descriptions oflanguages with metrical stress systems, this tonal event is called a pitch accent. Studies of American English intonation (e.g. Pierrehumbert 1980) and Australian English intonation (Fletcher and Harrington 1996) posit a relatively rich inventory of pitch accents, e.g. H* (high), L+H* (rising), L*+H ('scooped' rise), and L* (low), which are phonologically associated with one or more rhythmically prominent syllables in a word. Dyirbal, like English, appears to have a prominence-marking tonal event that can be linked with rhythmically prominent syllables in a word (King 1994). However, King found that pitch accents in Dyirbal narratives, which were the basis of her study, have essentially only one shape, LH*L.

The language under investigation in this chapter, Bininj Gun-wok (henceforth BGW), is a polysynthetic, non-configurational language of the Gunwinyguan family, spoken in western Arnhem Land (see Evans *1997a, 1997b,* 2003: sec. 2.5.2.2) and is genetically only distantly related to Dyirbal. Like the latter, however, it has conventionally been described as a stress language (Evans 1995; Fletcher and Evans 2000) with no lexical use of pitch. Its polysynthetic structure results in complex metrical stress rules (see Evans (1995,2003) and Bishop (2001, 2003: ch. 3) for a more detailed treatment than can be given here). Pitch accents can anchor to more than one metrically



												Embedded Verb							
-12	-11	-10	(-9)	(-8)	(-7)	(-6)	(-5)	(-4)	(-3)	(-2)	(-1)	(E -4)	(E -1)	(E -0)	(E+3)	0	(+1)	+2	(+3)
Tense	Subject	Object	Directional	Immediate	Miscellaneous	Benefactive	Miscellaneous	Generic IN	Body Part IN	Spatial/Num.	Comitative	Generic IN	Comitative	Stem [open]	Gerundive suffix	Verb root	Reflex/recip.	TAM	Case
	pronominal prefixes																		

FIGURE 12.1 Template structure of the verbal word in BGW.

strong syllable within the word: a single morphosyntactic word may carry two to three pitch accents.

Figure 12.1 illustrates the complexity of BGW morphosyntactic structure. The verbal word in BGW has the morphosyntactic template structure shown (adapted from Evans, forthcoming). The shaded 'slots' are obligatory, the parenthesized slots optional. Slots 6 and 7 are interchangeable in order. The relative linear order of the remaining slots is fixed, but it is never the case that all slots are filled at the one time. A Manyallaluk Mayali utterance, *bani-weleng-bepbe-marne-yaw-bu-rr-iny*, glossed as '3dualPast-then-separately-benefactive-baby-hit-each.other-perfective', and translatable as 'Then the two of them fought each other over the baby', illustrates the filling of slots (−12 (zero prefix), −11), (−7) (twice), (−6), (−4), (0), (+1), and (+2).

An example of double accentuation within a verbal morphosyntactic word (*ngarri-yauh-maknan*, 'we'll try looking at one more place') is given in Figure 12.2 below, and in a nominal word (*gun-marlaworr-dorreng*, 'with a leaf'), in Figure 12.3.

Previous intonational analyses of two of the dialects examined here, Gun-djeihmi (Fletcher and Evans 2000) and Manyallaluk Mayali (Bishop 2003; Bishop, Fletcher, and Evans 1999) showed that the language falls into the typology of having both prominence-lending pitch movements and edge-marking pitch movements. In other words, the intonational typology of these dialects appears to be closer to Bengali (Hayes and Lahiri 1991) than to Korean. In the following sections we will briefly describe the main intonational contours observed for BGW and outline rules of tune-text association that give rise to these specific intonational patterns. We will then describe the transcription conventions that have been developed to capture significant prosodic events.

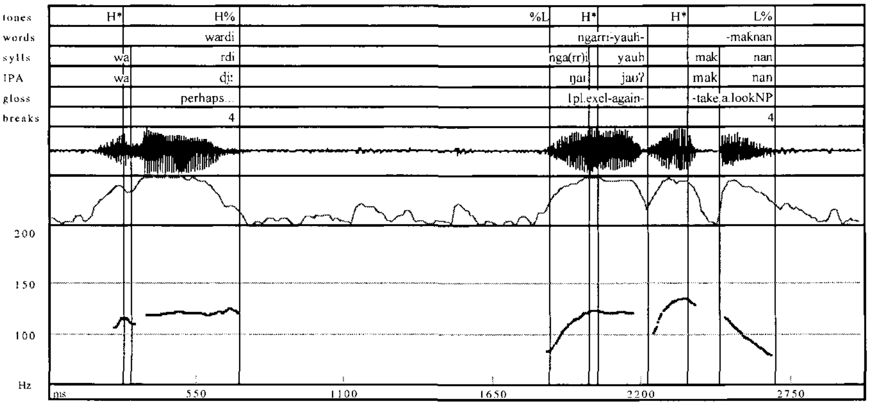


FIGURE 12.2 Hat pattern in Gun-djeihmi (across *ngarri-yauh-maknan*). Translation: *'We'll try looking at one more place'.*

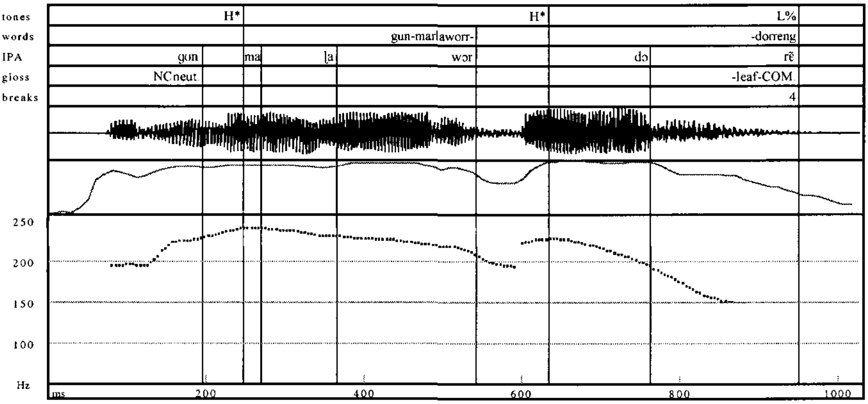


FIGURE 12.3 Hat pattern in Manyallaluk Mayali. Translation: *With a leaf.*

12.2. OVERVIEW OF THE CORPUS

All of the data presented in this chapter were recorded in the field by either the first author, Nicholas Evans, Murray Garde, or Peter Carroll as part of their fieldwork programmes on Northern Australian languages (Carroll 1995; Evans 2003; Garde 2003). The corpus consists of twelve narrative texts from 1.3 to 10 minutes in duration; one recording of an elicitation session and one of a conversation, mostly between two women, in Kuninjku; and two texts in Manyallaluk Mayali containing interrogative citation forms (twenty-six questions in total). The distribution of texts across the dialects is as follows: Gun-djeihmi (4), Kundedjnjenghmi (1), Manyallaluk Mayali (4), Kune (1), Kuninjku (5), and Kunwinjku (1). Of the Gun-djeihmi texts, three narratives were recorded from one male speaker and the remaining narrative text was produced by another male speaker. The Kundedjnjenghmi text was produced by one male speaker. Three of the Manyallaluk Mayali texts, including one set of citation forms, were spoken by one male speaker, with the remaining set of citation forms recorded from two female speakers. The Kune narrative was obtained from a female speaker, the Kunwinjku narrative from a male speaker, and the three Kuninjku narratives from a single male speaker. Intonation patterns in the latter texts were also found to occur in the text of a conversation between Kuninjku women. No gender-specific prosodic patterns are apparent in this corpus.

12.3. THE MAIN INTONATIONAL CONTOURS OF BININJ GUN-WOK: TUNES ASSOCIATED WITH DECLARATIVE, INTERROGATIVE, AND IMPERATIVE CONSTRUCTIONS

The main *declarative tune* consists of a 'pointed' or a 'flat' hat pattern, associated with an intonational phrase containing one or more pitch accented syllables, and a final fall to a low boundary. The hat pattern is found across all dialects. Figure 12.2 is an example from Gun-djeihmi and Figure 12.3 from Manyallaluk Mayali. General interpretations of this contour are similar to those associated with phrase-final falls in declarative utterances in other languages.

Another, less common, declarative tune consists of the hat pattern with a final rise at the right edge of the contour (Figures 12.4 and 12.5). The rise starts relatively low and ends at mid-level in a speaker's range. In all other respects, this contour resembles those illustrated in Figures 12.2 and 12.3. This contour is less frequent in the corpus than either of the other declarative tunes. In the Gun-djeihmi section of the corpus, for example, twenty-three out of a total of 221 intonational phrases are associated with this pattern.

The Fo range of the rise is generally not very large. Note, however, the extent of the rise in the Kune example (Figure 12.5), by contrast with the Manyallaluk Mayali example (Figure 12.4). Since the rise begins from a fairly low level in each of Figures 12.4 and 12.5, it does not appear to be the case that a distinct tune is involved. In the absence of evidence for the phonological

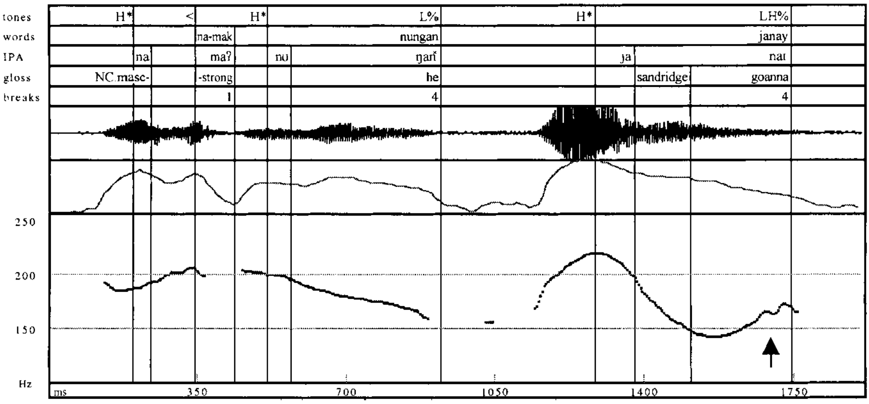


FIGURE 12.4 Rising boundary in Manyallaluk Mayali
Translation: *[But that (other) one], the strong one, sandridge goanna...*

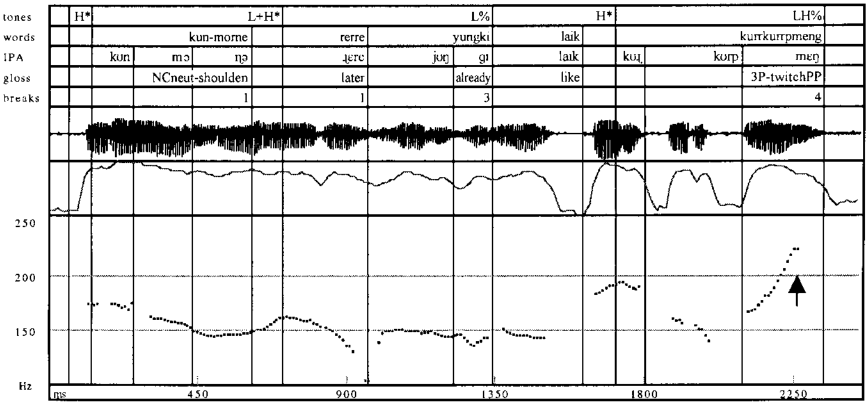


FIGURE 12.5 Rising boundary in Kune
Translation: *[Later, his father already knew that something had happened to his son] because his shoulder had been twitching.*[1]

status of the higher rising contour, the high rise in the second example is presently treated as a phonetic raising of the high target of the rise.

A notable feature of the corpus is indeed the absence of any evidence for a phonological high rising tune (of the kind English ToBI annotates H-H%). However, the corpus is limited in the genres it covers, and it is possible that expansion of the corpus to other genres (e.g. conversation) will require

[1] Translation from Evans, N. (2003), *Bininj Gun-wok: A Pan-Dialectal Grammar of Mayali, Kune and Kunwinjku* (Canberra: Pacific Linguistics).

expansion of the tonal inventory established to date. The third tune used in declarative utterances is a *high sustained contour* reaching from the final accented syllable to the phrase edge. There are two variants of this tune. The first is illustrated in Figure 12.6 for Kundedjnjenghmi and the second is illustrated in Figure 12.7 for Kuninjku. In the first variant, the stretch of pitch between the final accent in the phrase (also the first, in this example) and the phrase edge is sustained at a mid-high pitch level. In the second variant, referred to as the 'stylized' high sustained contour, sustained mid to high pitch is combined with stylized lengthening of the phrase-final syllable

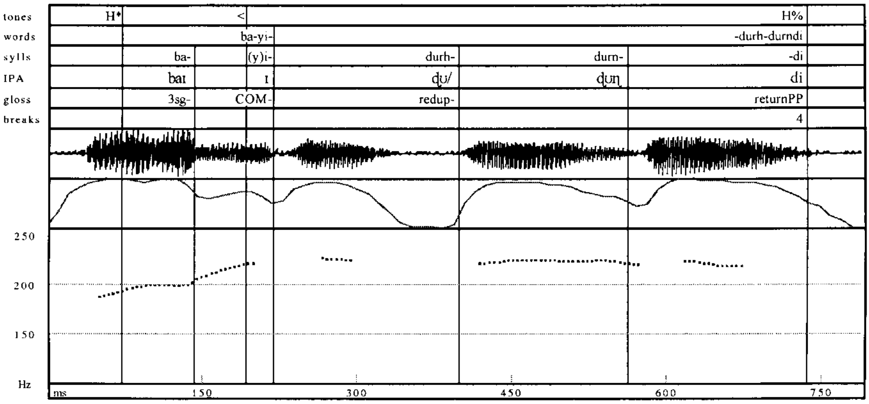


FIGURE 12.6 High sustained contour in Kundedjnjenghmi (non-stylized)
Translation: *He returned with it.*

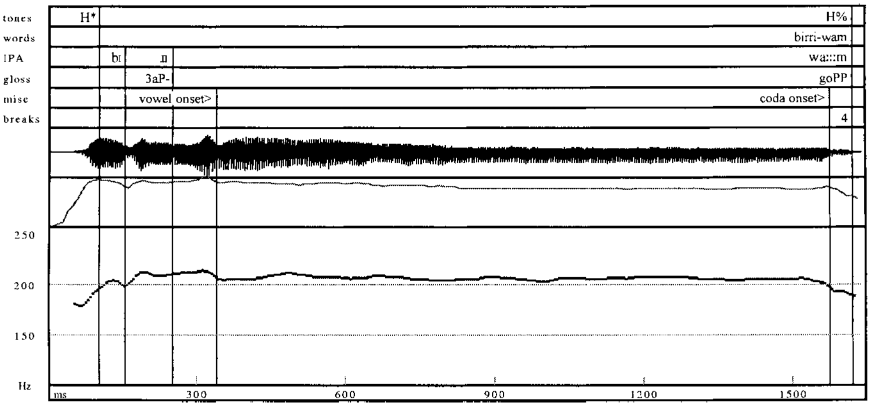


FIGURE 12.7 'Stylized' high sustained contour in Kuninjku (with vowel lengthening)
Translation: *They went along...*

nucleus. A coda consonant following the syllable nucleus is not stretched. Although this pattern is most frequently realized on the final vowel of verbal words, a nominal word following a verb may carry the vowel lengthening, indicating that the lengthening is a phrase-final edge effect rather than a kind of prosodic 'suffix' to the verbal word. The stylized sustained high contour type (Figure 12.7, above) is a strong feature of all dialects of BGW. An analogous contour has also been observed impressionistically among several languages in Australia, such as Alawa (Sharpe 1972), Nunggubuyu (Heath 1984), Iwaija (Birch 1999), and Wik-Mungkan (Sayers 1974). The similarity of the contour across these languages is apparently semantic as well as formal. The meaning cited by Sharpe, for example, is among the meanings of the contour in BGW: 'The pattern signifies continuous or prolonged action, motion, or state (according to the meaning of the verb)' (Sharpe 1972: 37).

Intriguingly, a somewhat similar phenomenon is recorded by Woodbury (1987) as occurring in two varieties of Central Alaskan Yupik (CAY), another polysynthetic language. The process of 'foot stretching' in CAY lengthens and raises the pitch of a foot-final segment (nucleus or coda consonant). The process may affect only the initial (leftmost) foot in the intonation phrase, or it may also affect subsequent feet. Though the domain and intonational position of the effect differ from Bininj Gun-wok, the iconic content overlaps: as in BGW, 'the degree of stretching... is entirely up to the speaker' and as the formal effect increases in magnitude, 'the intensification which it signals also increases' (Woodbury 1987: 716). However, in CAY the device also typically 'underscores the surprise value' of the information, which it does not appear to do in BGW. In BGW, it more often serves as a means of 'setting the scene', or dramatizing a continuous, but backgrounded action (signalled by the stylized sustained high intonation), which is then punctuated by a momentary action or event.

Our observations of the intonation of *interrogative* constructions are based on a small set of predominantly Wh-questions. The few examples of polar questions in the corpus display a phrase-final fall (Figure 12.8). The tune associated with Wh-interrogatives is a high pitch accent peak on the phrase-initial Wh-question word or a demonstrative adjacent to it, followed by a phrase-final fall-rise-fall pattern (Figure 12.9; Bishop 1999). The rise-fall section of the pattern is aligned with either the final syllable or the penultimate syllable of the phrase (see Section 12.5.5 below for a discussion of the transcription of this tune as involving an accent plus boundary tone sequence rather than a bi- or tritonal boundary tone).

Imperative utterances show a very similar pattern of final fall-rise-fall. The small set of imperative utterances in the corpus are predominantly by the same speakers as the interrogatives (Figure 12.10).

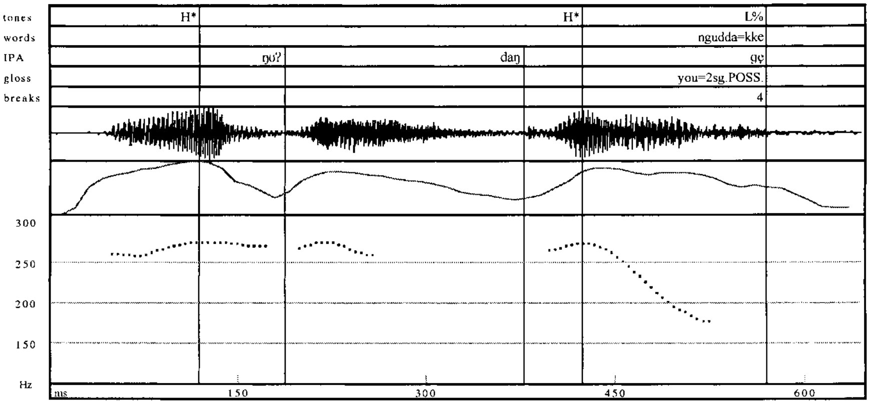


FIGURE 12.8 Polar question contour in Kuninjku
Translation: *'It's YOURS?'*

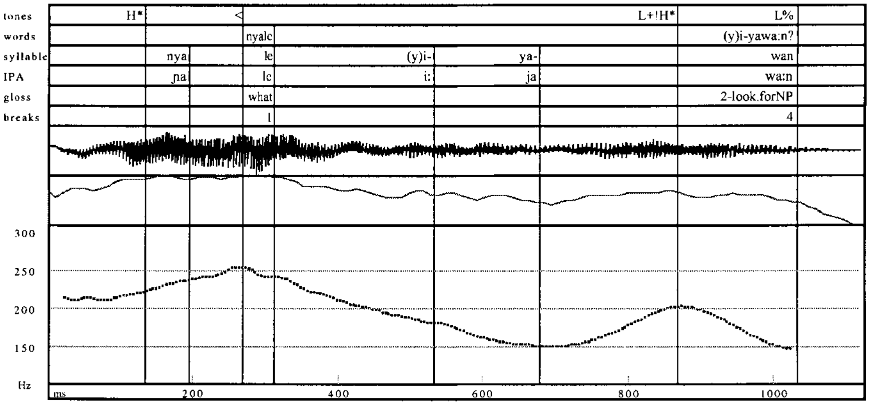


FIGURE 12.9 Wh-question contour in Manyallaluk Mayali
Translation: *What are you looking for?*

12.4. PROSODIC STRUCTURE AND TUNE-TEXT ASSOCIATION IN BININJ GUN-WOK

The segmental phonology of BGW provides few cues to prosodic domains of the kind described by Nespor and Vogel (1986) (see Section 12.5.7, 'Break Index 0', below). The principal cues are tonal and durational. The levels of prosodic structure found to be relevant to a description of BGW intonation thus far are the *foot*, the *phonological phrase*, the *intonational phrase*, and the *utterance.*

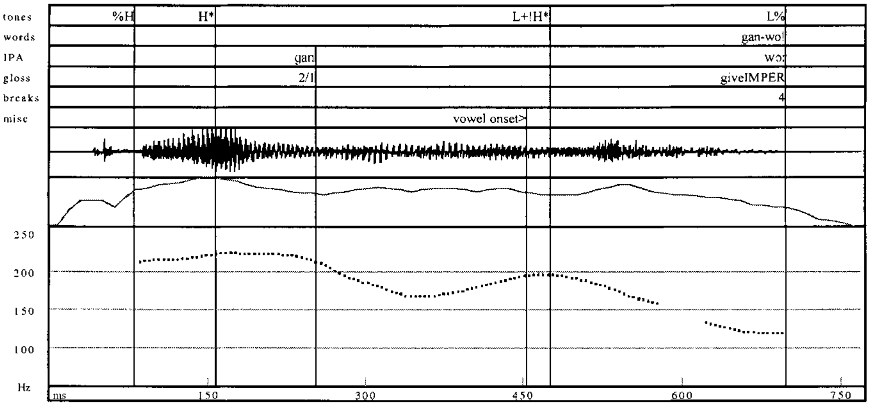


FIGURE 12.10 Imperative contour in Manyallaluk Mayali
Translation: *Give it to me!*

The *foot* is trochaic and unbounded in all the BGW dialects (Bishop 2001, 2003: 22, 119). Metrical structure is assigned on the basis of morphological structure; morphemes are generally isomorphic with feet. The principal exceptions are a small set of morphemes which conjugate the root for tense, mood, and aspect, and prosodically cohere with the root to form a single foot. Evidence for the unboundedness of feet comes from tri- and quadrisyllabic monomorphemic nominal words, which bear stress on the initial syllable only, and, in unemphatic speech, do not carry a pitch accent on any other syllable:

1 (*
0 * * * *
gorlomomo 'fresh water crocodile' (Manyallaluk Mayali dialect)

not:

1 (* (*
0 * * * *
***gor**lo**mo**mo

This stress pattern in monomorphemic words contrasts with the bimorphemic tri- and quadrisyllabic words. These regularly bear a pitch accent on each foot, which provides evidence of their bipedal metrical structure.

1 (* (*
0 * * * *
detjmak-**du**ninj lit. hero-proper 'a real brave guy'
(Manyallaluk Mayali dialect)

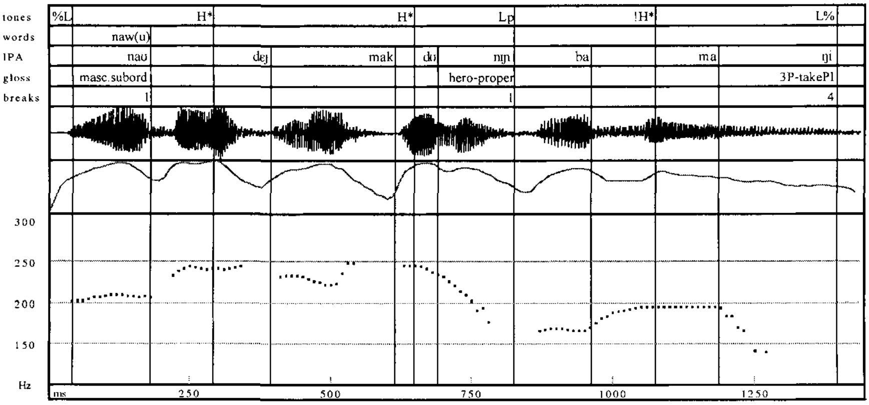


Figure 12.11 Double accented, bimorphemic nominal word in Manyallaluk Mayali
Translation: *That real brave guy would take it.*

The morphological basis of metrical structure assignment often leads to adjacent stresses and even adjacent pitch accents. The prosodic system of BGW is unusual in showing a high tolerance for adjacent prosodic heads: there is no evidence of either stress- or accent-clash effects (Bishop 2001). Since the head of any foot in the morphological word provides a potential landing site for pitch accent, the morphological word in BGW may bear one, two, or exceptionally, three accents. We hear these as generally having similar perceptual prominence.

In the Kuninjku dialect, the *phonological phrase* is a level of phrasing that is tonally marked (with a low tone) at its right edge (see Figure 12.12, below). The phonological phrase is immediately dominated by the intonational phrase within the prosodic hierarchy postulated for BGW. Grammatically, the phonological phrase corresponds to the maximum level of lexical projection (VP, NP). As yet, no evidence has been gathered as to whether this level is also tonally marked in the other dialects, though, impressionistically, there is a similar phenomenon in the Kunwinjku and Manyallaluk Mayali dialects (see Figures 12.11, 12.14, and 12.19, in which the low phonological phrase edge tone is tentatively marked).[2] In Kuninjku, relative prominence relations (downstep and upstep) hold between the final accents in adjacent phonological phrases. Detailed phonetic and phonological argumentation for

[2] An alternative analysis of the Low tone in these three examples is L+H*. Further evidence (of the kind illustrated in Figure 12.12 for Kuninjku, in which there is a clear low target marking the edge of the word *bi-rrulubom*) is needed to determine which analysis is appropriate.

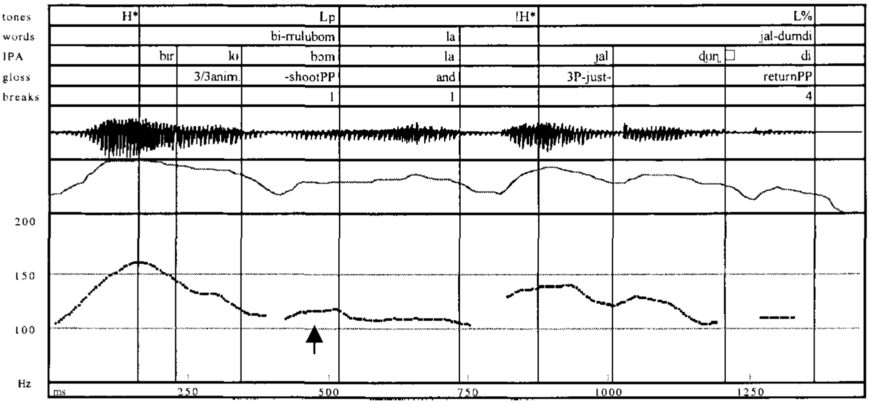


Figure 12.12 Low Phonological Phrase boundary tone (Lp) in Kuninjku
Translation: *He shot/speared him and just came back.*

the tonal marking of the phonological phrase level in Kuninjku is given in Bishop (2003: ch. 6). No final lengthening is associated with the phonological phrase level in that dialect.

The *intonational phrase* is defined on the basis of three characteristics: the relative prominence of pitch accents, boundary tone/s, and pitch reset. The intonational phrase is the domain within which the relative prominence relationships of downstep and upstep are constituted. Among the accents in the phrase, one accent tends to sound more prominent than the others: this is the 'head' or 'nucleus' of the intonational phrase. A single boundary tone or bitonal sequence associates with the right edge of the phrase, and is realized on the final syllable or two of the phrase. At the left edge of the phrase a new choice of pitch range is made, and downstep/upstep are blocked. The left edge may also be optionally marked by an initial boundary tone, %L or %H (see Section 12.5.6). There is no clear evidence of phonetic boundary-associated lengthening at the level of the intonational phrase.

A sequence of intonational phrases constitutes an *utterance*. There are two principal characteristics of the utterance: the potential for final lowering and substantial pause. A low intonational phrase boundary tone at the right edge of an utterance may be phonetically lowered relative to preceding low boundaries, producing the effect of final lowering. Not all intonation phrase-final falls show this extra low pitch. Also, the contours that display this feature do not always indicate that the speaker is 'done', and as yet we have found no consistent pragmatic or discourse function associated with this local lowering (Fletcher and Evans 2000). The end of the utterance is generally followed by more substantial pause than occurs utterance-medially. There are no additional boundary tones associated with the edge of the utterance. However, the final two syllables of the utterance generally undergo phonetic lengthening as a correlate of the boundary (Fletcher and Evans 2000; Bishop 2003: 355–9).

12.5. TRANSCRIBING BININJ GUN-WOK INTONATION

12.5.1. *Overview of BGW-ToBI*

In this section we outline the BGW tonal inventory and transcription conventions adopted for the six dialects of Bininj Gun-wok. Our transcription system is closely based on the Tones and Break Indices system developed for General American English (Pitrelli *et al.* 1994) and several other languages examined in this volume, for example Australian English (Fletcher and Harrington 1996), Korean (Beckman and Jun 1996), Japanese (Venditti 1997), and German (Grice *et al.* 1996). We will show how the same labelling conventions can be applied to all six dialects but will highlight how the few dialectal differences observed to date can be accounted for. A table summarising these conventions can be found in the Appendix.

Due to the morphological complexity of BGW we have added one further tier, a morphological gloss tier, to the usual four tiers of the classic ToBI model. All data are currently labelled using a modified version of 'Transcriber' developed for English ToBI. 'Transcriber' is an ESPS/Xwaves shell that has linked 'menu' files for each xlabel field. In some cases, these fields have been modified from the original English ToBI menus to account for Bininj Gun-wok-specific prosodic structure. For example, the tone menu has been modified to include specific diacritics for 'upstep' (^, as in ^H*) and 'Final Lowering' ('Final_Lo'). The minimum signal requirements for basic labelling include the acoustic waveform, fundamental frequency curve, and linked xlabel files. In some cases, spectrograms are also generated along with RMS amplitude traces to facilitate stressed syllable location and word level transcription. The five 'core' tiers currently included in the BGW-ToBI transcription are:

(1) a word tier
(2) a gloss tier
(3) a tone tier
(4) a break-index tier
(5) a miscellaneous tier.

12.5.2. *Word tier*

The orthographic word tier in BGW-ToBI is similar to the orthographic tier in other ToBI systems. Words are transcribed using the conventional orthography developed by linguists in consultation with the community. Each word label is linked to the final segment of the word. Spectrograms are used where necessary to facilitate location of these right edges. Any hesitations or pauses are included in the miscellaneous tier.

Bininj Gun-wok is a highly polysynthetic language, so in many cases words are extremely long. As a general rule, hyphens have been used to indicate morphological division (refer to the Appendix for a key to the morphological glosses used in the illustrations to this paper). In the prosodic system of BGW, the morphosyntactic word generally corresponds to the phonological word. However we have observed certain instances of disjuncture between the morphological and phonological word in a closely related language, Dalabon, with certain prefixes appending to preceding phonological words. This will be the subject of further study.

12.5.3. *The morphological gloss tier*

A tier containing a morphological gloss is included in the conventions. This is essential because (1) labellers are not native speakers of the language; and (2) the study involves multiple dialects. The requirement of a gloss tier responds to the need for the researchers to understand the semantic structure of the utterances analysed. This understanding is of particular importance in the labelling of words and word groupings for which the cues to prosodic structure alone may be ambiguous.

The morphological gloss tier (illustrated in Figure 12.13 below) has the additional advantage of providing the basis for a cross-dialectal comparison of the relationship between prosodic and morphological structure. In all dialects, the rules for the construction of prosodic feet access the morphological structure. However, constraints upon the construction of feet and the position of primary stress within the prosodic word vary between dialects in ways which are still in the process of being formulated (Bishop 2001).

12.5.4. *Tone tier*

There are two basic tone types included in the labelling conventions for the tonal tier—pitch accents and boundary tones. These conventions are

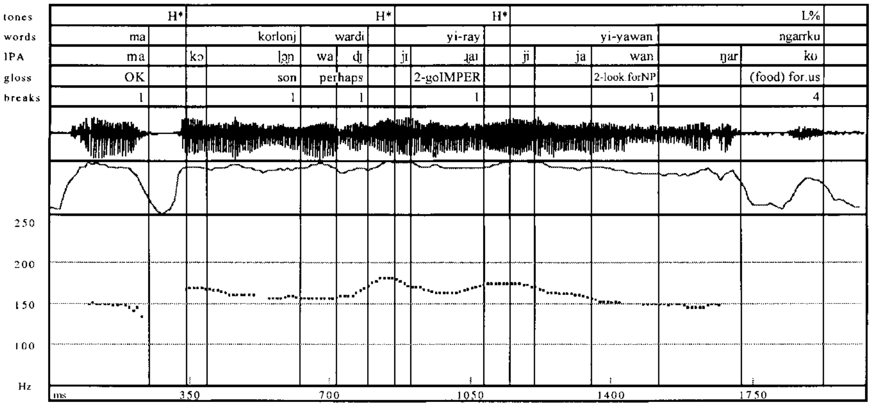


FIGURE 12.13 Word and morphological gloss tiers (Kune)
Translation: *OK, my son, you go and look for something [for us to eat].*

appropriate for all dialects in question, although we will discuss potential cross-dialectal realization differences at the end of this paper. Pitch accents are associated with stressed syllables and boundary tones with the right edge of phonological and intonational phrases. Rules for alignment of these tone types are outlined in Sections 12.5.5 and 12.5.6 below.

The inventory of tunes in BGW is relatively sparse compared with that of prototypical intonation languages such as Dutch and English.[3] There are five pitch accent types, four of which are monotonal (high, delayed high, lowered (downstepped) high, and raised (upstepped) high) and one bitonal (low-rising). There is one type of phonological phrase boundary tone (low), and three types of intonation phrase boundary tone (low, high, and (low-) rising). The pitch accent types are discussed in Section 12.5.5, and the boundary tone types in 12.5.6.

12.5.5. *Types of pitch accents: (transcribed with *)*

Simple high (H*) This is the main accent type in BGW. H* is realized as a rise from the onset of the accented syllable to a peak within that syllable,

[3] It may be that BGW uses other aspects of the prosody to produce inferential meanings which are created in languages such as Dutch and English by the paradigmatic choice of tune. These aspects might include, for example, phrasing and dephrasing, pitch range modifications, and possibly modifications of tune-text alignment (Bishop 2003).

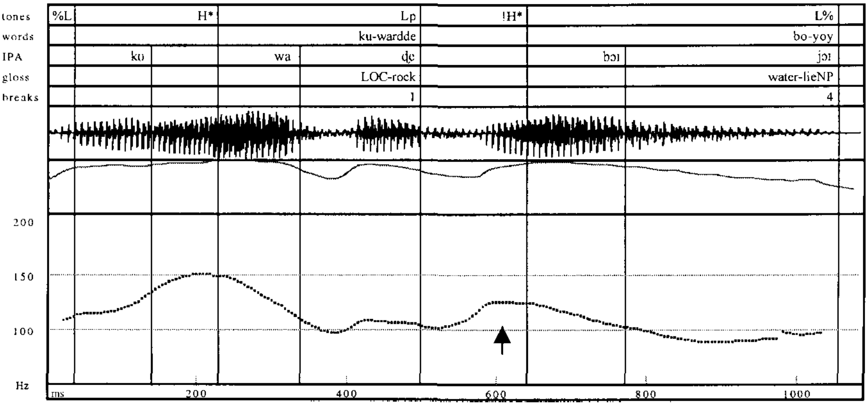


FIGURE 12.14 Downstepped high accent (!H*) in Kunwinjku
Translation: *Water lay in the cave.*

usually aligned late in the syllable rhyme. The tone target is generally scaled within the mid-upper part of the speaker's Fo range.

Delayed high[4] (H* <) The delayed high accent is realized as a rise from the onset of the accented syllable to a peak in the post-stress syllable (see Figures 12.4, 12.6, and 12.9 above). Following the ToBI conventions used for Korean and English, we label the highest Fo point with an additional diacritic ' < ' to indicate 'late' peak (the angled bracket points back to the stressed syllable with which the peak is phonologically associated). The peak is never delayed beyond the post-stress syllable.

Simple downstepped[5] (!H*) The tone target is lowered relative to a preceding high tone target within the intonation phrase. The alignment of the downstepped tone is similar to that of H* simple accents. The preceding H target can be a simple H* followed by a fall to a low phonological phrase boundary tone, an L+H* rising accent (described above) or an intonational phrase initial high boundary tone (%H; see above), which creates a 'high prehead' extending up to the first accent in the phrase.

Terraced downstepping contours of the kind noted in many other languages (Ladd 1996) are frequently found in all of the BGW dialects. Examples of this contour from Kunwinjku (Figure 12.14) and Kundedjnjenghmi (Figure 12.15) are given below. An upstepping sequence of accents occurs less frequently,

[4] This accent type is referred to as the 'late peak' accent in Bishop (1999, 2000, 2003: 249ff).

[5] In Bishop (2003), !H* and ^H* are treated as *modifications* of underlying H*. Also, L% is reanalysed as Lp-%, representing the association of the low phonological phrase boundary tone (Lp) to the intonation phrase boundary. However, these distinctions, and the reasons for them, cannot be developed in the present brief treatment of the inventory.

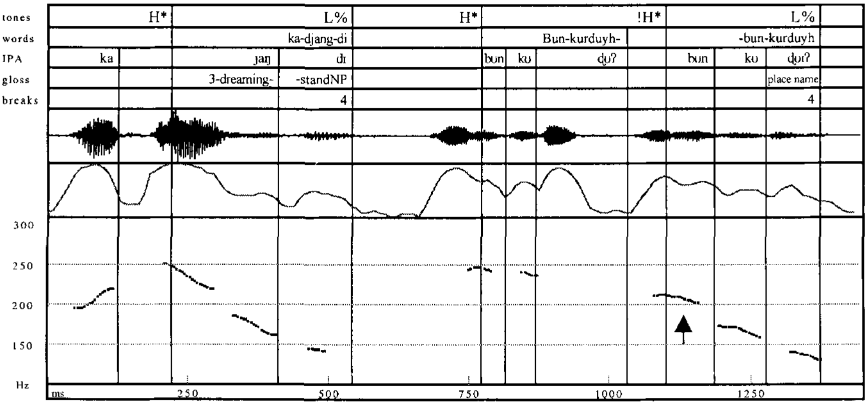


FIGURE 12.15 Downstepped high accent (!H*) in Kundedjnjenghmi
Translation: *There's a dreaming at Bunkurduyh-Bunkurduyh.*

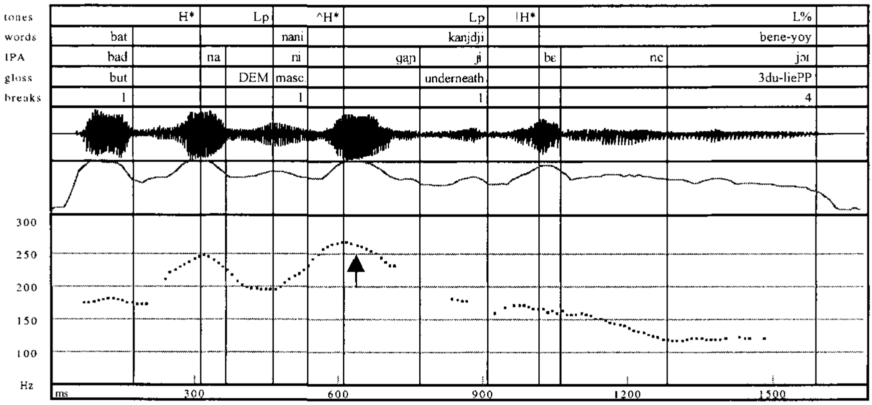


FIGURE 12.16 Upstepped high accent (^H*) in Kuninjku
Translation: *But those two were lying underneath.*

and may be followed by a downstepped accent within the intonation phrase (Figure 12.16).

Simple upstepped (^H*) The tone target is raised relative to a preceding high accentual tone target within the intonation phrase. It is aligned in the same manner as H* and !H*. So far, this accent has only been attested in the Kuninjku data (Figure 12.16).

Bitonal (low rising) (L+H*) This pitch accent consists of a high tone preceded by a rise from the low part of a speaker's pitch range. The rise is generally observed through the stressed syllable. In some cases the L is realized on a preceding syllable if the preceding material is highly sonorant.

It is necessary to distinguish this bitonal accent from a tonal sequence consisting of either an intonation phrase-initial L tone (transcribed %L; see below), or a phonological phrase-final L tone (Lp), followed by H*. An %L H* sequence is particularly common in the Gun-djeihmi dialect. There are two ways in which L+H* accents can be distinguished from such a sequence. There is usually tight temporal coupling between the low and high tone in the case of the bitonal accent. In the case of the %L H* sequence, the L is clearly anchored at the left phrase edge and can occur at some distance from the first H* accent.

L+H* is also only labelled where the preceding L tone could not be accounted for as a phonological phrase-final low tone. In this respect, it is significant that we have observed word-medial instances of L+H* accents, where we cannot account for the L tone as a boundary tone associated with a preceding phrase edge (Figures 12.9 and 12.10). The rise in L+H* tends to be somewhat sharp and the H target is generally realized in the upper part of a speaker's range. However, in a sequence of L+H* tones the second H* may be downstepped, and is labelled L+!H* accordingly (this is the case in Figures 12.9 and 12.10 above).

L+H* is a less frequent accent type in BGW. It may be dialect-specific in its distribution, as it does not occur in the data from the Kuninjku dialect analysed to date. In Gun-djeihmi narratives, the accent is usually employed to signal emphasis or narrow focus (Fletcher and Evans 1998). Similarly, in Kune and Manyallaluk Mayali, it sounds a more emphatic accent than a simple H* pitch accent.

A particular use of the L+H* tone in Manyallaluk Mayali is in Wh-questions and imperative utterances (Figures 12.9 and 12.10). The phrase-final fall-rise-fall pattern which is associated with these utterances was originally analysed as L* HL%, with the HL% sequences being realized as a rise-fall in the absence of a phonological upstep rule in the language. However, re-analysis of the alignment of the first L tone and the relatively strong auditory prominence of the H in the HL sequence indicates an analysis of L+ (!)H* L% is a more appropriate description of the tone pattern and its prosodic structure.

12.5.6. *Types of boundary tone*

One level of boundary tone is presently distinguished for all dialects: the intonational phrase level. A detailed analysis of Kuninjku (Bishop 2003: ch. 6) has also provided evidence for a tonally marked phonological phrase level in that dialect.

Phrasal tones are labelled at the end of each intonational phrase. The tone labels are generally aligned with the right edge of the last word in the phrase, as labelled on the orthographic tier. A feature of the boundary tone labels in BGW is that there is no phonological upstep rule for boundary configurations, in contrast with American and Standard Southern British English, or German (Grice *et al.* Ch. 13 this volume). That is, there is no evidence for an H-H% boundary tone sequence, denoting a high-rising tone, but only H%, used in BGW-ToBI to denote a level high tone, and LH%, used to denote a low-rising tone. The balance between the most frequent boundary tone types, L% and H%, varies considerably across texts. Although L% tones generally predominate, the percentage of L% boundary tones across all texts varied from 42 per cent to 93 per cent. H% tones constituted between 7 per cent to 58 per cent of all boundary tones, while LH% tones were rare, at between 0 per cent and 3 per cent of boundary tones.

Low phonological phrase boundary tone (Lp) The phonological phrase boundary tone is realized as a fall from the final high pitch accent in the phonological phrase to a low target aligned with the penultimate or final syllable of the phrase (see Figures 12.12, 12.14, 12.16). The phonological phrase is frequently isomorphic with a single morphosyntactic nominal or verbal word, but occasionally extends to two words (Bishop 2003: ch. 7).

Low intonational phrase boundary tone (L%) This event is generally realized as a fall from the last pitch accent of a phrase reaching a low target near the baseline of a speaker's pitch range. The low target aligns with the penultimate or final syllable of the phrase (see Figures 12.5, 12.12, 12.15, 12.16). It is sometimes problematic to locate this tone clearly on the Fo curve if the speaker is elderly, as there is often a high level of creak which accompanies low boundary tones. The L% tone may be additionally lowered in utterance-final position by the final lowering modification (see Section 12.5.7, 'Break Index 4'), conveying a stronger sense of finality than the unmodified L%.

High intonational phrase boundary tone (H%) This is realized as a sustained high pitch or a slight rise from the last H* accent in a phrase. (Refer to Figures 12.6 and 12.7 above.) There is evidently no rule of upstep: in a sequence of H* H%, the H% is not necessarily realized higher in the speaker's pitch range than the preceding accent. This is similar to H% boundary tones in Glasgow English ToBI (Mayo 1996), which also undergo no upstep. H% is a common boundary configuration in narratives for all dialects. It is principally found utterance-medially, in descriptions of sequential actions, and in the initial phrase of a disjunction; the final boundary tone in such sequences is L%.

Low-rising intonational phrase boundary tone (LH%) This boundary configuration phonetically resembles the 'continuation rise' of the English ToBI system. It generally involves a rise from the lower part of a speaker's range to a mid-level range, with both the low and the high target aligned with the final syllable of the phrase (refer to Figures 12.4 and 12.5). The range of meanings of LH% in BGW is unclear, as there are few examples in the corpus. However, one use of LH% would seem to be to solicit the hearer's attention to a new or newly re-introduced referent, as in the phrase *but nawu... (L%) na-mak nungan janay (LH%)* '[but that (other) one]... the strong one, sandridge goanna...'.

Initial high boundary tone (%H) This represents a clear target in the high part of the speaker's pitch range which is not necessarily associated with a phonologically stressed syllable. %H extends a level high plateau across any speech material preceding the first pitch accented syllable in the phrase (see Figure 12.17: the tone ending the preceding intonational phrase in this example is L%). This tone, and the initial low initial boundary tone described below, are similar to the 'pre-head' of the British School of intonation. %H is observed in all dialects.

Initial low boundary tone (%L) This represents a target at the onset of an intonational phrase that is distinctly low in the speaker's range (as opposed to a default, mid-level onset), and does not align closely with a following high

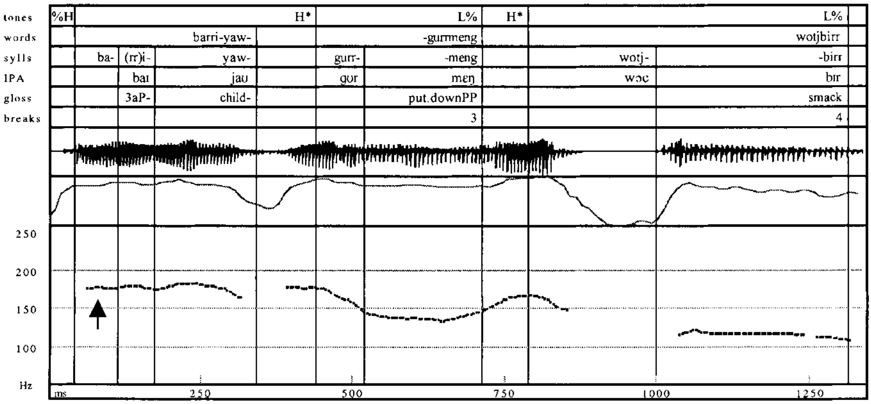


FIGURE 12.17 High Initial Intonational Phrase boundary tone (%H) in Manyallaluk Mayali
Translation: *They put the child down, smacked it.*

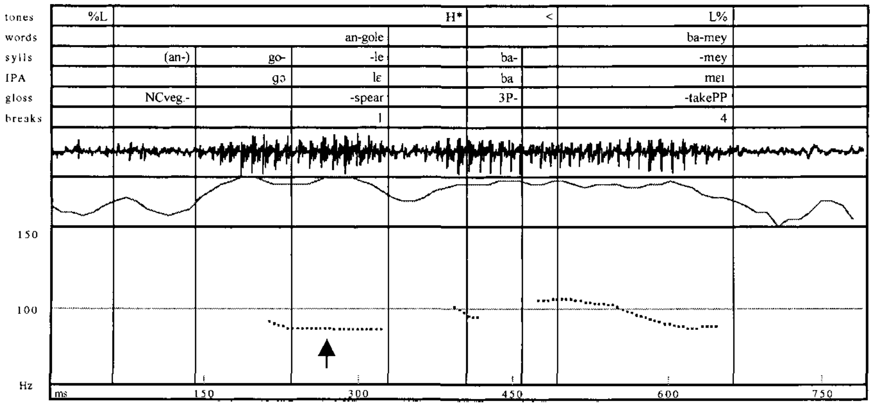


FIGURE 12.18 Low Initial Intonational Phrase boundary tone (%L) in Gun-djeihmi Translation: *He took a spear.*

pitch accent, as the Low target does in the bitonal accent (L+(!)H*) (see Figures 12.9 and 12.10). When the first word in the phrase is unaccented, %L produces a low 'pre-head', a stretch of speech (in Figure 12.18, the word (*an*)-*gole*), with which a 'floating' Low tone target is associated. It is principally required for the description of tunes in Gun-djeihmi.

12.5.7. *Break indices*

The system of break indices enables the labelling of perceived degrees of prosodic juncture. Some levels of prosodic juncture are not tonally marked, and therefore not captured in the tonal tier, but may nonetheless be of relevance to the discourse and grammatical structures of languages. The break indices system may also be used to index the additional prosodic correlates of boundaries which *are* tonally marked, such as pause.

Four break index labels are adopted in the present analysis of Bininj Gun-wok: 0, 1, 3, and 4. The content of these labels is outlined below.

The break index labels for BGW map directly to two of the levels of the hierarchy of prosodic structure described in Section 12.4 above (the intonational phrase, BI 3 and the utterance, BI 4) and distinguish another (the phonological word and phonological phrase have the same level of perceived juncture, BI 1).

Break Index 0: Break Index 0 is used where segments at a morphosyntactic word boundary are phonetically elided, and/or the position of a word boundary is reanalysed in fast speech. That is, the 'underlying' prosodic level of Break Index 0 is the level usually indicated by Break Index 1 (see below). It does not mark a separate level in the prosodic structure of BGW. In our corpus, Break Index 0 does not occur more frequently at some kinds of syntactic juncture than at others; nor are there *systematic* segmental phonological processes, such as liaison or epenthesis, with which the break index is associated. Therefore, for our purposes, Break Index 0 is considered to have the status of a phonetic label in a narrow transcription. In an utterance labelled '*ma* 0 *wurd* 1', for example, a glottal stop at the end of the '*ma*' is elided and the word boundary erased. A less transparent example is '*gun-dulk* 0 *yi-rratj(je)* 1' (/gʊndʊɫk˺ ## jɪraɪc/, pronounced [gʊndʊɫ ## kɪraɪc], in which there is both elision of the initial glide /j/ and reanalysis of the word juncture. In the example illustrated in Figure 12.19 below, '*ngal(i)* 0 *(ng)albu*', the final vowel /ɪ/ in the first word and the velar nasal onset of the second are elided.

Break Index 0 should only be used where it is not possible to discern the original word boundary. In all other cases, 1 or a higher level is preferable. Thus, the frequency of occurrence of Break Index 0 in our corpus, even in fast speech, is relatively low.

Break Index 1: Break Index 1 is the default, marking the minimal degree of juncture between a pair of morphosyntactic words in phrase-medial position. There is a distinct perception of the final and initial segments of the words, usually in the absence of physical pause. Examples of Break Index 1 are in Figures 12.19 and 12.20.

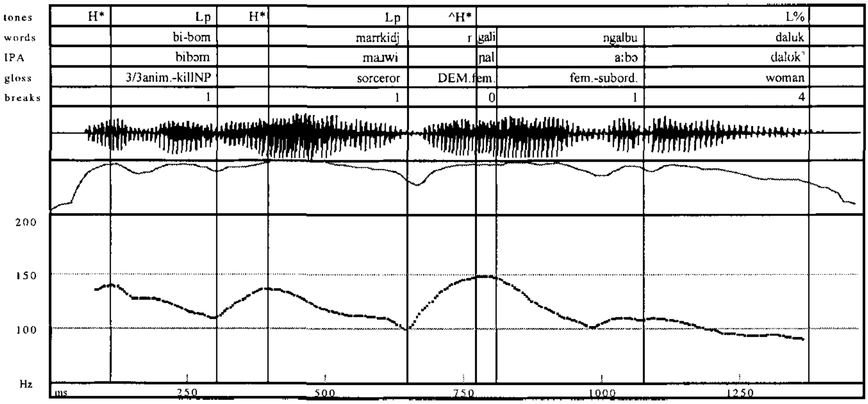


FIGURE 12.19 Break Indices 0, 1, and 4 in Kunwinjku
Translation: *The clever man killed her, that woman.*

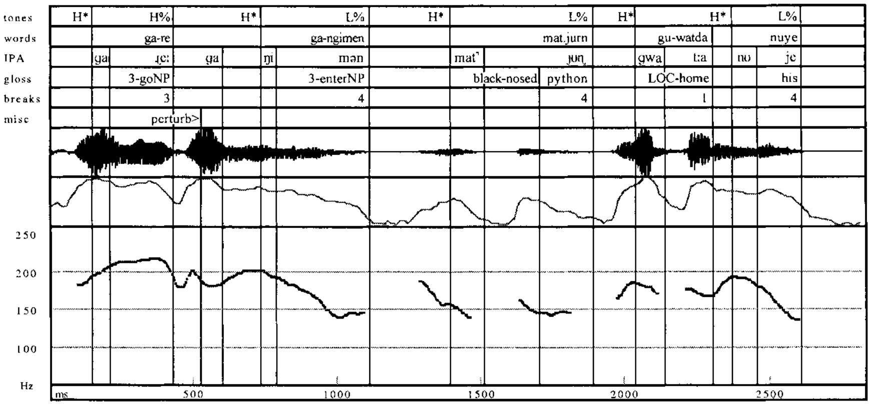


FIGURE 12.20 Break indices 1, 3, and 4 in Manyallaluk Mayali
Translation: *The black-nosed python goes along, (then) into his hole.*

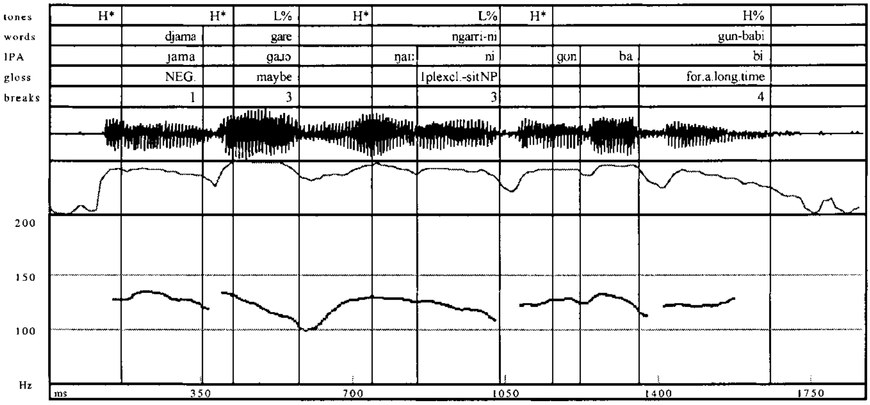


FIGURE 12.21 Break Indices 1, 3, and 4 in Gun-djeihmi
Translation: *Maybe we won't stay here long.*

Break Index 3: Break Index 3 is generally labelled at the right edge of an utterance-non-final intonational phrase (see Figures 12.20 and 12.21). There is at least one strong prominence per phrase, and a boundary tone realized at the right edge of the phrase. At the level of Break Indices 3 and 4, the cues to juncture are a combination of a 'cumulative' cue occurring within the prosodic unit (the relative tonal prominence of the head) and 'demarcative' cues occurring at its edge (pause, lengthening, and boundary tone) (cf. Swerts *et al.* 1994). Clear pitch reset in the following phrase is a strong, though not a consistent, cue. Phrase-final lengthening (or 'virtual pause') is not a systematic cue to Break Index 3. Impressionistically, Break Index 3 is also not associated with long pause; any pause tends to be no greater than 100 milliseconds following this juncture. Since substantial pausing tends to coincide with the percept of finality, Break Index 4 has been reserved for contours displaying one or both of these cues.

Break Index 4: Break Index 4 is used to mark an utterance-final intonational phrase juncture (see Figures 12.19, 12.20, and 12.21 above). This juncture resembles a Break Index 3 juncture tonally and with regard to its prominence features, but has *one or more* additional features: substantial pause, a sense of finality (associated with final lowering), or stylization (refer to Section 12.3).

Each of these features alone is sufficient to warrant labelling as Break Index 4. Finality is judged on the basis of whether, were the utterance to be synthetically cut off at the boundary, a listener would judge the speaker to have finished what they had to say (cf. Swerts *et al.* 1994). Though long pause is one correlate of Break Index 4, and a sense of finality another, it is not necessarily the case that long pause is a *cue* to finality in Mayali. A doctoral study of the rhetorical structure of Kunwinjku narratives by Carroll, for example, finds that '[p]ause units are grouped into tone units with the final pause unit having a falling intonation pattern' (Carroll 1995: 118). Carroll defines 'pause units' by the presence of a pause lasting at least 200 milliseconds, and 'tone units' by the fall to the baseline, which causes the percept of finality. It is clear that substantial pause is not associated uniquely with finality, but is used, at least in narrative, as a rhetorical and structural device (Carroll 1995). The extent to which pause is correlated with the level of the utterance (Break Index 4) may differ across languages and genres.

12.5.8. *Miscellaneous tier*

The miscellaneous tier is primarily used to note a variety of disruptions to the speech signal or the pitch trace, such as creak or phrase-final devoicing at low pitch; pitch doubling and halving, and microprosodic perturbations of the trace; disfluencies, hesitations and interruptions or turn-taking by another speaker (who may be identified in the miscellaneous tier for the purpose of discourse study); laughter; and various noises (children, dogs). It may also be used to mark silences (<sil sil>) for the purpose of studying the use of pause, and, where relevant, to indicate a stress or accent that is unusually positioned within the word (e.g. 'final syllable accented').

12.5.9. *Optional customized tiers: the syllable tier*

The syllable tier may serve as an important aid for research on languages in which the morphological word is regularly polysyllabic. Together with the morphological gloss tier, a syllable tier enables the determination of which morphemes are susceptible to elision, or fusion with neighbouring morphemes, and under what prosodic conditions. The syllable tier was also found to be useful for the study of tonal alignment in BGW (cf. Bishop *et al.* 1999).

12.5.10. *Additional features of our labelling system*

There are two additional features associated with the labelling of pitch range.

HiFo: we have retained the same convention as English ToBI to mark the highest Fo value associated with a high accent in each intonational phrase. Only Fo targets associated with H* accents are labelled. Segmental perturbations are taken into account when labelling HiFo.

Final_Lo: we have added this optional label to our tone menu to represent the 'extra' lowering of pitch that can occur at the end of an utterance-final intonational phrase. With elderly speakers, there is also a high level of creak accompanying these 'extra low' boundary tones. This label is particularly relevant for our work on intonation and discourse (cf. Fletcher and Evans 2000).

12.6. CROSS-DIALECT AND CROSS-LANGUAGE DIFFERENCES

Differences among the dialects for which data has been analysed do not appear to be substantial. The few differences we have posited are predominantly of the 'systemic' type, that is: 'differences in the inventory of phonologically distinct tune types, irrespective of semantic differences' (Ladd 1996: 119, following Wells (1982)). For example, the initial low intonational boundary tone %L frequently occurs in Gun-djeihmi, but does not appear to feature in the inventory of Kuninjku. On the other hand, upstepping accents within an intonational phrase are only recorded for Kuninjku. These may be simply gaps in the data; further data in each dialect, from a larger number of speakers and a more extensive range of genres is required to address this question.

We would like to conclude with some general comments on the prosodic patterns observed in our Bininj Gun-wok corpus. Speakers of the Bininj Gun-wok dialects examined here use intonational parameters such as pitch range reset, local downstep, and final lowering extensively to provide certain kinds of discourse effects: initiating a new topic, introducing new participants, closing an old topic, and so forth. Our preliminary observations suggest that in this respect, BGW does not differ from other intonational languages.

Given the polysynthetic nature of this language, we speculate that there may be fewer instances of word-level prosodic juncture (Break Index 1) in Bininj Gun-wok compared to more morphologically isolating languages. Many morphosyntactic words in BGW constitute BI 3 or BI 4 units. These words range from one syllable to upwards of seven syllables in length. A survey of accent frequency across Manyallaluk Mayali, Kune, Kunwinjku, and Gun-djeihmi texts showed a ratio of pitch accents to words ranging between 0.71 (a Manyallaluk Mayali text) to 0.92 (Kunwinjku). This is a high rate of accentuation, which may reflect (a) the absence of deaccentuation as a pragmatic device in Bininj Gun-wok and (b) the paucity of unaccented 'function' or non-'content' words in this language. The ratio of accents to words is also raised by the presence of double accentuation in certain kinds of morphological words, such as complex verbal constructions.

A survey of boundary tone frequency across Manyallaluk Mayali, Kune, and Kunwinjku texts showed a ratio of 1.5 to 2.1 words per boundary tone (Break Index 3 or 4). The high frequency of boundary tones reflects the fact that much information in the clause is incorporated into the verb in BGW (see Figure 12.1), reducing the use of free nominal and adverbial words to instances where further specification of information (that is otherwise simply indexed on the verb) is needed. There is therefore a close mapping between a single clause (minimally, a verb with a pronominal prefix and inflectional suffixes) and a single intonational break (BI 3 or 4). An examination of two Manyallaluk Mayali narratives showed 91 per cent of clauses in one narrative (thirty-one out of thirty-four clauses) and 86 per cent of clauses in another (sixty-nine out of eighty clauses) map onto single intonational phrases. It would be interesting to compare our findings to other polysynthetic languages to see if similar trends are evident.

APPENDIX

Key to morphological glosses used in the text

Verbal prefixes

1, 2, 3	First, second, third person
sg., pl., a, du	Singular, plural, augmented (3+), dual

anim.	Higher animate
P	Past
3/3	Third person subject affecting third person object
excl./incl.	Exclusive/inclusive
BENE.	Benefactive
COM.	Comitative
redup.	Reduplicant morpheme

Verbal suffixes

NP	Non-past
PI	Past imperfective
PP	Past perfective
IMPER.	Imperative modality

Nominal prefixes

NCmasc./fem./veg./neut	Noun class marker (masculine, feminine, vegetable, and neuter classes)
LOC.	Locative prefix

Nominal suffixes

LOC.	Locative suffix
COM.	Comitative suffix
POSS.	Possessive suffix

Other

masc.subord./fem. subord.	Subordinating conjunction *-bu* with masculine/feminine gender agreement
NEG.	negative particle
DEM.masc./fem	Demonstrative with masculine/feminine gender agreement

Summary of BGW_ToBI labels

H* *Simple high accent*: marked on a stressed syllable. Indicates an accent rising from the onset of the stressed syllable to an Fo peak on that syllable.

H*< *Delayed high accent*: marked on a stressed syllable. Indicates an accent rising from the onset of the stressed syllable to an Fo peak on the poststress syllable.

!H* *Simple downstepped accent*: marked on an Fo peak that is lowered relative to a preceding high accent peak.

^H* *Simple upstepped accent*: marked on an Fo peak that is raised relative to a preceding high accent peak.

L+H* *Bitonal (low rising) accent*: marked on an Fo peak preceded by a rise from the lower part of the speaker's range.

Lp	*Low phonological phrase boundary tone*: marked at the right edge of a phonological phrase.
L%	*Low intonational phrase boundary tone*: marked at the right edge of an intonational phrase.
H%	*High intonational phrase boundary tone*: marked at the right edge of an intonational phrase displaying sustained high pitch after a final Fo peak in the phrase.
LH%	*Low-rising intonational phrase boundary tone*: marked at the right edge of an intonational phrase.
%L	*Initial low intonational phrase boundary tone*: marked at the left edge of an intonational phrase. Indicates an onset at a distinctly low level.
%H	*Initial high intonational phrase boundary tone*: marked at the left edge of an intonational phrase. Indicates an onset at a distinctly high level, but not associated with a high Fo accent peak.
0	*Break index*: *zero disjuncture*: indicates it is not possible to discern a word boundary, typically in fast and casual speech.
1	*Break index*: *weak disjuncture*: typical of word- and phonological phrase-level boundaries.
3	*Break index*: *medium disjuncture*: typical of utterance-medial intonational phrase boundaries.
4	*Break index*: *strong disjuncture*: typical of utterance-final boundaries.
HiFo	*Highest Fo*: marked at the highest Fo value associated with a high accent peak in each intonational phrase.
Final_Lo	*Final lowering*: marked at the right edge of an utterance displaying 'extra' lowering of pitch, signalling finality.

Note: the IPA font used in the Figures and in the body of the text is SIL-Doulos Regular (12).

REFERENCES

BECKMAN, M. E. (1996), 'The Parsing of Prosody', *Language and Cognitive Processes*, 11: 17–67.

——, and JUN, S.-A. (1996), 'K-ToBI (Korean ToBI) Labeling Conventions, Version 2', ms, Ohio State University and University of California, Los Angeles.

——, and PIERREHUMBERT, J. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook*, 3: 255–309.

BIRCH, B. (1999), 'Prominence Patterns in Iwaija', B.A. (Hons.) dissertation (University of Melbourne, Australia).

BISHOP, J. (1999), 'The When, Where and How of the Intonational Structure of Wh-questions in Manyallaluk Mayali', *Melbourne Papers in Linguistics*, 18 (Department of Linguistics & Applied Linguistics, University of Melbourne).

Bishop, J. (2000), 'Prosodic Context and the Timing of Fo Peaks: Late Peak Alignment in Bininj Gun-wok', paper presented at the Australian Linguistics Society conference, Melbourne, Australia, 7–9 July.

—— (2001), 'Metrical Structure, "Primary Stress" and Intonational Pitch Accent in Kuninjku', paper presented at the Australian Linguistics Society conference, Canberra, Australia, 27–30 Sept.

—— (2003), 'Aspects of Intonation and Prosody in Bininj Gun-wok: An Autosegmental-Metrical Analysis', Ph.D. dissertation (University of Melbourne, Australia).

——, Fletcher, J., and Evans, N. (1999), 'Tonal Alignment in Mayali', in *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco, CA), 2371–4.

Bruce, G. (1977), *Swedish Word Accents in Sentence Perspective* (Lund, Sweden: CWK Gleerup).

—— (this volume Ch. 15), 'Intonational Prominence in Varieties of Swedish Revisited'.

——, Elert, C., Englstrand, O., and Wretling, P. (1999), 'Phonetics and Phonology of the Swedish Dialects', in *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco, CA), 321–4.

Carroll, P. (1995), 'The Old People Told Us: Verbal Art in Western Arnhem Land', Ph.D. dissertation (University of Queensland).

Evans, N. (1995), 'Current Issues in the Phonology of Australian Languages', in J. Goldsmith (ed.), *The Handbook of Phonological Theory* (Cambridge, MA: Blackwells), 723–61.

—— (1997*a*), 'Role or Cast? Noun Incorporation and Complex Predicates in Mayali', in A. Alsina, J. Bresnan, and P. Sells (eds.), *Complex Predicates* (Stanford: CSLI), 397–430.

—— (1997*b*), 'Head Classes and Agreement Classes in the Mayali Dialect Chain', in M. Harvey and N. Reid (eds.), *Nominal Classification in Aboriginal Australia* (Amsterdam: John Benjamins), 105–47.

—— (2003), *Bininj Gun-wok. A Pan-dialectal Grammar of Mayali, Kunwinjku and Kune* (Canberra: Pacific Linguistics).

Fant, G., and Kruckenberg, A. (1993), 'Towards an Integrated View of Stress Correlates', *Proceedings of the ESCA Workshop on Prosody* (Lund: Lund University).

Fletcher, J., and Evans, N. (1998), 'Intonational Categories in Mayali', paper presented at Laboratory Phonology VI, York, 2–4 July.

——, —— (2000), 'Intonational Downtrends in Mayali', *Australian Journal of Linguistics*, 20/1: 23–38.

——, and Harrington, J. (1996), 'Accentual-Prominence-Enhancing Strategies in Australian English', in P. McCormack and A. Russell (eds.), in *Proceedings of the VIth Australian International Conference on Speech Science and Speech Technology* (Canberra: Australian Speech Science and Technology Association), 577–80.

Garde, M. (2003), 'Topics in Kuninjku Ethnography of Speaking', Ph.D. dissertation (University of Queensland).

Gordon, M. K. (this volume Ch. 11), 'Intonational Phonology of Chickasaw'.

Grice, M., Baumann, S., and Benzmüller, R. (this volume Ch. 3), 'German Intonation in Autosegmental-Metrical Phonology'.

——, Reyelt, R., Benzmüller, R., Maher, J., and Batliner, A. (1996), 'Consistency in Transcription and Labelling of German Intonation with GTOBI', in *Proceedings of the 4th International Conference on Spoken Language Processing* (ICSLP: Philadelphia), 1716–19.

Gussenhoven, C., and Vliet, P. van der (1999), 'The Phonology of Tone and Intonation in the Dutch dialect of Venlo', *Journal of Linguistics*, 35: 199–235.

Halliday, M. A. K. (1967), *Intonation and Grammar in British English* (The Hague: Mouton).

Hayes, B., and Lahiri, A. (1991), 'Bengali Intonational Phonology', *Natural Language and Linguistic Theory*, 9: 47–96.

Heath, J. (1984), *Functional Grammar of Nunggubuyu* (Canberra, Australia: Australian Institute of Aboriginal Studies).

Hermann, R. (1997), 'Syntactically-Governed Accentuation in Balinese', *OSU Working Papers in Linguistics*, 50: 679–99.

Jun, S.-A. (1993), 'The Phonetics and Phonology of Korean Prosody', Ph.D. dissertation (Ohio State University). [Published in 1996, New York: Garland.]

——, and Fougeron, C. (1995). The Accentual Phrase and the Prosodic Structure of French, in *Proceedings of the thirteenth International Congress of Phonetic Sciences* (Stockholm), 2: 722–5.

King, H. (1992), 'Dyirbal Intonation', in J. Ingram and J. Pittam (eds.), *Proceedings of the Fourth Australian International Conference on Speech Science and Speech Technology*, 597–601.

—— (1994), *The Declarative Intonation of Dyirbal: An Acoustic Analysis*, MA dissertation (Australian National University).

—— (1999), 'High Onset Pitch Accents? The Case of Dyirbal and Warlpiri', in *Proceedings of the XIVth International Congress of Phonetic Sciences* (San Francisco, CA), 2403–6.

Ladd, D. R. (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

Mayo, C. (1996), 'Prosodic Transcription of Glasgow English: An Evaluation Study of Gla-ToBI', M.Sc. dissertation (University of Edinburgh).

Nespor, M., and Vogel, I. (1986), *Prosodic Phonology* (Dordrecht: Foris).

Pierrehumbert, J. (1980), 'The Phonetics and Phonology of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

Pitrelli, J. F., Beckman, M., and Hirschberg, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToBI Framework', *Proceedings, 1994 International Conference on Spoken Language Processing*, 1: 123–6.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991), 'The Use of Prosody in Syntactic Disambiguation', *Journal of the Acoustical Society of America*, 90/6: 2956–70.

Sayers, B. (1974), *Interpenetration of Stress and Pitch in Wik-Mungkan Grammar and Phonology (Part 1)* (Summer Institute of Linguistics).

Sharpe, M. (1972), *Alawa Phonology and Grammar* (Canberra: Australian Institute of Aboriginal Studies).

Swerts, M., Bouwhuis, D. G., and Collier, R. (1994), 'Melodic Cues to Perceived Finality of Utterances', *Journal of the Acoustical Society of America*, 96/4: 2064–75.

Vaissière, J. (1995), 'Phonetic Explanations for Cross-Linguistic Prosodic Similarities', *Phonetica*, 52: 123–30.

Venditti, J. (1997), 'Japanese ToBI Labelling Guidelines', *Ohio State University Working Papers in Linguistics*, 50: 127–62.

Woodbury, A. (1987), 'Meaningful Phonological Processes: A Consideration of Central Alaskan Yupik Eskimo Prosody', *Language*, 63/4: 689–740.

13

Strategies for Intonation Labelling across Varieties of Italian

Martine Grice, Mariapaola D'Imperio, Michelina Savino, and Cinzia Avesani

13.1. INTRODUCTION

In this chapter we examine the intonation of a number of varieties of Italian. Since there is no agreement as to what constitutes 'Standard Italian intonation' (see Lepschy and Lepschy 1977; Galli de' Paratesi 1985), we shall not make it an aim of this paper to define such a standard. Instead, we take four geographically defined varieties: from the South, those spoken in Naples, Bari, and Palermo; and from Central Italy the variety spoken in Florence, with a view to establishing a common framework for annotating the phenomena which have so far been studied in these varieties. Although they have all been analysed in an autosegmental-metrical framework, we shall see that there are considerable differences in how this framework is used, and, as might be expected, differences in the phenomena selected for detailed study. Part of our task will be to attempt to piece together this fragmentary picture. A common annotation will facilitate the exchange of data in both variety-specific and in cross-variety studies. It will make it more straightforward to evaluate how far evidence from one variety can be used in support of a phonological analysis of other varieties. It will also enable us to analyse a more diverse set of speech styles than is possible within a smaller more restricted study. In fact, the speech data referred to in the different accounts in this paper range from spontaneous or semi-spontaneous dialogues to specially produced read or scripted speech in the laboratory.

We would like to thank Barbara Gili Fivela for fruitful discussion and advice during the writing of this chapter. We would also like to thank our co-authors of the oral presentation, Giovanna Marotta, Patrizia Sorianello, and Mary Caputo, for sharing their data with us and for participating in stimulating discussions during the development of this annotation scheme. Finally, we would like to thank Jörg Peters and an anonymous reviewer for helpful comments on an earlier draft.

Italian is a free-stress language, with predominantly penultimate, but also final, antepenultimate and even pre-antepenultimate stress (Lepschy and Lepschy 1977; D'Imperio and Rosenthall 1999). In the classification of languages according to their rhythmic type, it has been suggested that Italian is syllable-timed, along with other Romance languages (Bertinetto 1981; Farnetani and Busà 1999; Ramus *et al.* 1999). However, not all varieties have the same rhythmic properties (for Tuscan and Northern Italian, see Vayra *et al.* 1984, 1987; Farnetani and Kori 1986; for Southern varieties, see Romito and Trumper 1989). It has even been argued by Romito and Trumper (1989) that certain Southern varieties tend more towards stress-timing. We concentrate on tonal aspects of annotation, the *To* (Tones) part of ToBI, as these have been studied in more detail than rhythmic or prosodic phenomena, especially in relation to database annotation. This means that the *BI* (Break Index) part of ToBI is only dealt with superficially.

In the following sections we describe those features of intonational structure shared by all of the varieties examined, and provide an account of those which have been described in one or more varieties, pointing out where similar phenomena should be sought in other varieties. Section 13.2 deals with pitch accents, especially those typical of statements and yes-no questions. It also deals briefly with downstep. Section 13.3 discusses the evidence for two levels of intonational phrasing and proposes how differing degrees of juncture can be transcribed using Break Indices. Then in Section 13.4 we return to pitch accents in order to discuss the notion of nuclear pitch accent, providing a definition appropriate for varieties of Italian. Section 13.5 takes a look at a number of postnuclear prominences and investigates which of them can be analysed as 'phrase accents'. Finally, in Section 13.6 we propose a way of annotating the partial realization, or truncation, of phrase-final pitch contours.

13.2. PITCH ACCENTS

All of the varieties described here have both falling (H+L) and rising (L+H) nuclear pitch accents, as well as monotonal ones (L* and H*). However, the inventories are not identical. Table 13.1 gives a summary of the nuclear pitch accents as transcribed in each account (see Section 13.4 for a discussion of nuclearity). Prenuclear accents will not be discussed here in any detail since they have not been the focus of any of the accounts available. It appears that a reduced set of accents occur prenuclearly, with a predominance of H* (see Figure 13.7 and Figure 13.8 below) and L+H* (Figure 13.1). Although L* has also been attested (Figure 13.5), its transcription is less straightforward. If we

Table 13.1 Nuclear pitch accents and their uses across the four varieties of Italian

	Neap.	Bari	Palermo	Flor.
Declarative broad focus	H+L*	H+L*	H+L*	H+L*
Declarative contrastive focus	L+H*	H*+L	H*+L	H*
Yes/no question BF & CF	L*+H	L+H*	L*+H	H*
Continuation	L*	L*	L+H*	

compare Figure 13.5 and Figure 13.6 below, which represent equivalent utterances in Neapolitan and Bari respectively, we can observe that where Neapolitan has a prenuclear L* accent, Bari has no accent at all. This type of alternation across the transcription systems needs further investigation.

13.2.1. *Statements*

We can observe from Table 13.1 that the nuclear accents occurring in broad focus declaratives are analysed in the same way in all varieties.

There is no such consensus when dealing with contrastive narrow focus declaratives. However, despite the different analyses, there are indications that the realization of narrow focus, especially if it is contrastive, does not differ considerably from one variety to another. Neapolitan narrow focus declaratives were initially transcribed as having a H*+L tone, like Bari and Palermo. The reason for selecting this particular pitch accent in Neapolitan, as in the other varieties, was based on the fact that there is a sharp fall on the accented syllable, completed either on the accented syllable itself or in the following syllable. The Neapolitan reanalysis of this accent as a L+H* is based on close examination of the focus constituent medial valley, which is analysed as the leading tone of a rising (L+H) pitch accent (D'Imperio 1999). In fact, differently from questions, statement focus constituents made of more than one word present a constituent medial fall, which immediately follows the initial rise, as is shown in Figure 13.1. In this and the following examples, the focused constituent is enclosed in square brackets.[1]

What was found is that the medial Fo minimum consistently presents the same value, independently of the amount of segmental material separating the first and second accent in the constituent. Through regression line fitting it was also shown that, by increasing the number of syllables intervening between the two accents, the slope of the contour from the preceding peak to

[1] The 'n' label is used to denote the nuclear pitch accent. See Section 13.4.

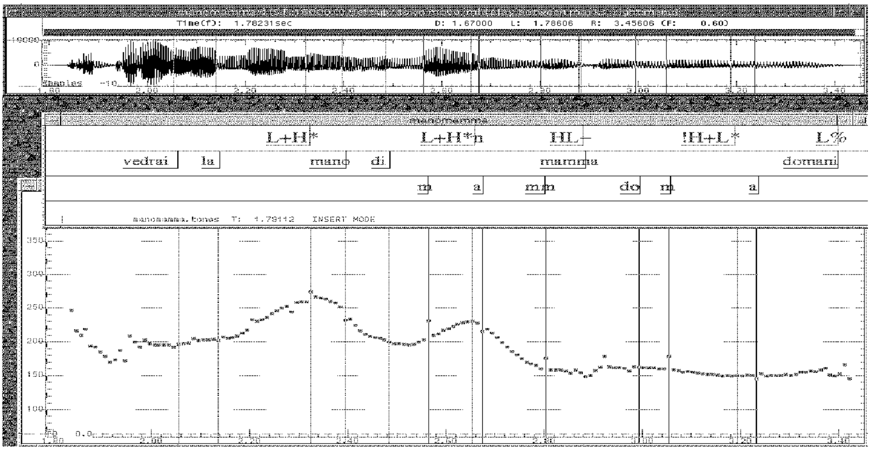


Figure 13.1 The constituent medial fall analysed as the L component of the narrow focus L+H* accent in Neapolitan. 'Vedrai la [MAno di MAMma] domani' (*You'll see [Mom's hand] tomorrow*).

the medial low becomes shallower. This points to the existence of an actual L target, structurally belonging to the nuclear accent in the narrow focus constituent. It must also be noted that such a result was essential for discarding the plausible hypothesis of a simple 'sagging' interpolation between two subsequent H peaks within the constituent (cf. Pierrehumbert 1980). As a consequence of the new L+H* analysis, the constituent final fall cannot be attributed to the pitch accent structure, unless one allows for a tritonal LHL pitch accent, which is dispreferred. We return to the final fall in Section 13.5.

Florentine uses neither H*+L nor L+H* for narrow focus declaratives but rather H* followed by an L- intermediate phrase boundary. The decision to account for the fall with a L- phrase accent is based on results of an experiment where the end of the fall was found to be at the end of the focused word (Avesani and Vayra 2000). This analysis is like Neapolitan in that the fall is not accounted for by the accent alone, but it is also similar to the other analyses in that there is no leading L tone. An examination within each variety of the preaccentual valley along the lines of D'Imperio (1999) and of the fall along the lines of Avesani and Vayra (2000) would provide us with a clearer picture.

The situation is complicated further: in Bari and Palermo, non-contrastive narrow focus declaratives can be produced with the same pitch accent as broad focus declaratives, thus leading to ambiguity as to narrow or broad focus, as is the case for a number of languages including English (Ladd 1996).

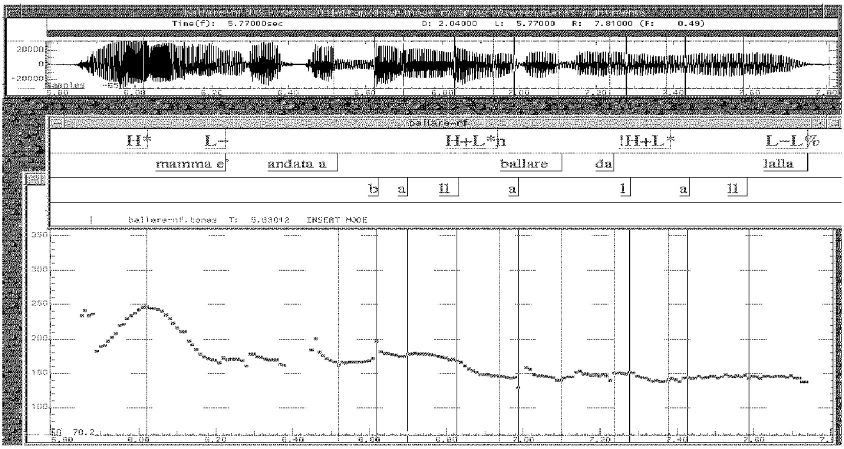


FIGURE 13.2 'MAMma [è andata a balLAre] da LALla' (*Mom [went to dance] at Lalla's*). Example of narrow focus (non contrastive) declarative in Bari Italian.

An example of a non-contrastive narrow focus declarative is given in Figure 13.2. It could be conceived of as an answer to the question 'Cosa è andata a fare mamma da Lalla?' ('What did Mom go to do at Lalla's?'). This is different from contrastive narrow focus, an example of which is shown below in Figure 13.9, which would imply a question such as 'È andata a *mangiare* mamma da Lalla?' ('Did Mom go to *eat* at Lalla's?'). Note in Figure 13.2 that the second H+L* accent is prefixed by an exclamation mark (!). This label indicates that the accent is downstepped, a discussion of which is deferred to Section 13.2.3.[2]

13.2.2. *Yes/no questions*

There is a good deal more variation in the accents used in yes-no questions. Before dealing with the intonation patterns, it is important to point out that what is described as yes-no question intonation is usually that of information-seeking yes-no questions (referred to as queries, see Carletta *et al.*, 1997). In the case of Bari, the same intonation pattern is used for tentative confirmation-seeking yes-no questions (tentative checks). A check is referred to as tentative

[2] Also, note that the first H+L*, though not marked as downstepped, presents a lower Fo target than the preceding H*. This is because H+L* possesses an 'intrinsically' downstepped quality which is retained even when the accent is not preceded by prenuclear material. This is akin to the H*+L pitch accent in Greek ToBI (Arvaniti and Baltazani this volume Ch. 4) which, owing to the reduced pitch range within which it is always produced, had originally been transcribed as a downstepped accent.

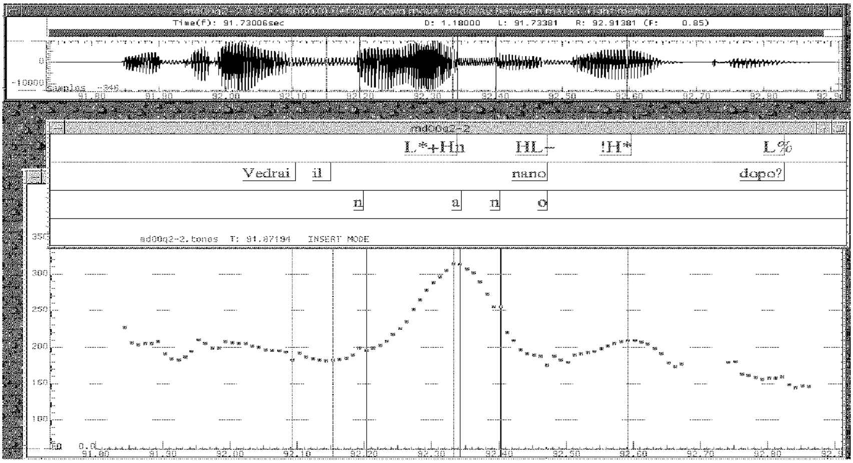
if the speaker's confidence as to the correctness of inferred material—i.e. that information is old—is very low (Grice and Savino 1997). We shall see below that a more confident check is produced with a different intonation pattern (for a systematic description of intonation contours in Bari Italian queries and checks see Savino 1997).

We can see from Table 13.1 that information-seeking yes-no questions can take L*+H, L+H* and H* as their nuclear pitch accent. The primary cue to interrogation in the Southern varieties is the pitch accent: L+H* in Bari Italian and L*+H in Palermo and Neapolitan, after which there is usually a final fall. A final rise represented as a high boundary tone constitutes an optional stylistic variant in Bari and Palermo. In Florentine, the pitch accent alone is not sufficient to signal unambiguously that an utterance is a question. To do this, a final rise, represented as a high boundary tone, is obligatory. It is important to note here that rising pitch accents (without necessarily a high boundary tone) have been reported in yes-no questions in a number of Central and Northern varieties (Endo and Bertinetto 1997; Marotta and Sorianello 1999; Gili Fivela 2003). That is, they are not restricted to the Southern varieties. An overview of the most common yes-no question patterns in each variety examined in this paper, concentrating on the nuclear pitch accent and phrasal tones, is given in Table 13.2.

Although the three Southern varieties all use a rising pitch accent to mark questions, the type of pitch accent is not the same: Neapolitan and Palermo have a L*+H pitch accent, where the pitch is low on the accented syllable and rises up from it. The rise may or may not be completed within the accented syllable, depending on factors such as syllable structure and the vicinity of the accented syllable to the phrase boundary. Compare for instance the timing of two instances of L*+H in Neapolitan: in L*+H on 'na' of 'nano' in Figure 13.3 the rise is completed within the stressed vowel, whereas in 'bel' of 'bella' in Figure 13.4 the target for the LH rise is realized only at the end of the word. We return to a discussion of the pitch contour after the rise in Section 13.5.1.

Table 13.2 Typical intonation patterns in information-seeking yes-no questions

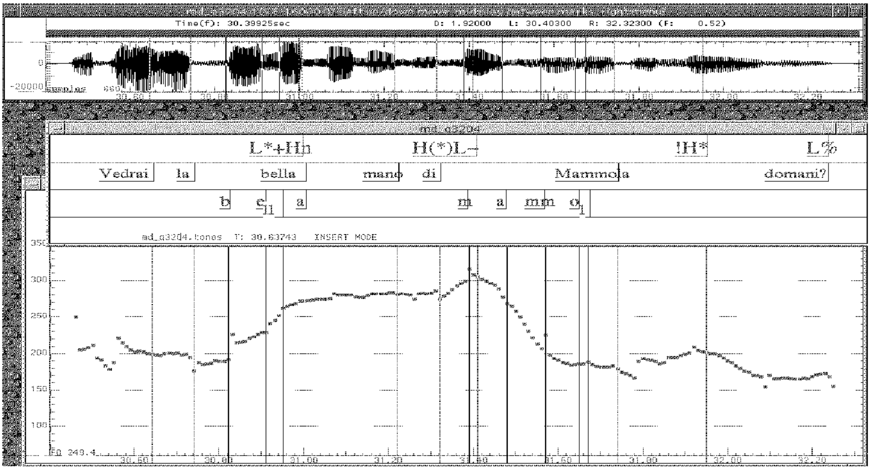
	Question tune	
	Nuclear accent	Phrasal tones
Neapolitan	L*+H	HL- L%
Bari	L+H*	L-L% or L-H%
Palermo	L*+H	L-L% or L-H%
Florentine	H*	L-H%



Time(f): 91.73006sec D: 1.18000 L: 91.73381 R: 92.91381 (F: 0.85)

md00q2-2.tones T: 91.87194 INSERT MODE

FIGURE 13.3 Neapolitan 'Vedrai il [NAno] DOpo?' (*Will you see the [dwarf] afterwards?*).



Time(f): 30.39925sec D: 1.92000 L: 30.40300 R: 32.32300 (F: 0.52)

md_q3204.tones T: 30.63743 INSERT MODE

FIGURE 13.4 Neapolitan 'Vedrai [la BELla mano di MAMmola] doMAni?' (*Will you see [Mammola's beautiful hand] tomorrow?*).

Since there are many cases, both in Palermo and in Neapolitan, where both the L and the H occur within the accented syllable, alignment facts alone would have been insufficient to decide upon the starredness of the L tone. In both varieties the decision is the consequence of a contrast within the phonological system between the question accent and another rising pitch accent. In Palermo, the L*+H contrasts with an earlier timed rise in a type of

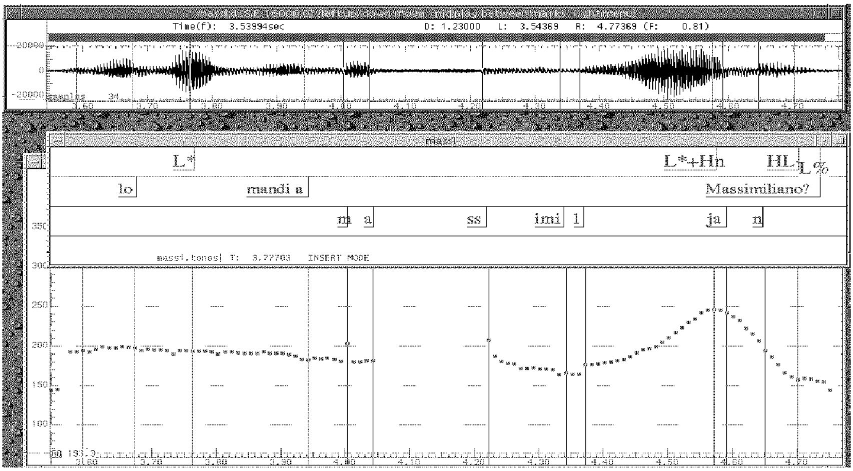


FIGURE 13.5 Neapolitan: 'Lo mandi a [MassimiLIAno]?' (*Will you send it to [Maximillian]?*).

continuation contour which takes the L+H* label. Similarly, the reanalysis in Neapolitan of the declarative narrow focus contour as having a rising L+H* pitch accent motivated the notational contrast and thus led to the use of L*+H in questions.

Bari Italian is different from Palermo and Neapolitan in that the peak is reached around the middle of the accented syllable. A comparison with Neapolitan, where the peak is reached at the end of the accented syllable can be made by observing Figure 13.5 and Figure 13.6. The question arises as to whether on functional grounds Bari questions should be annotated using the same pitch accent as the other Southern varieties,[3] the L*+H pitch accent. Our initial decision here is negative; until an earlier-timed rising pitch accent has been found which would contrast with the question accents, we advocate keeping the label as surface-oriented as possible, taking association of a tone to a syllable to be indicated by at least an approximate synchronization in time between the two entities.

Although it is clear that Italian uses intonation to distinguish between yes-no questions and statements, the transcriptions given above suggest that narrow focus declaratives in Neapolitan have a similar contour to yes-no questions in Bari. This leads immediately to the question of whether

[3] The suggestion that the same pitch accent should be used across varieties was also made by Marotta (2000), although she argues that the interrogative rising pitch accents should be transcribed as L+H*, an option which is not available to us for the Neapolitan and Palermo data on the grounds of contrastivity given above.

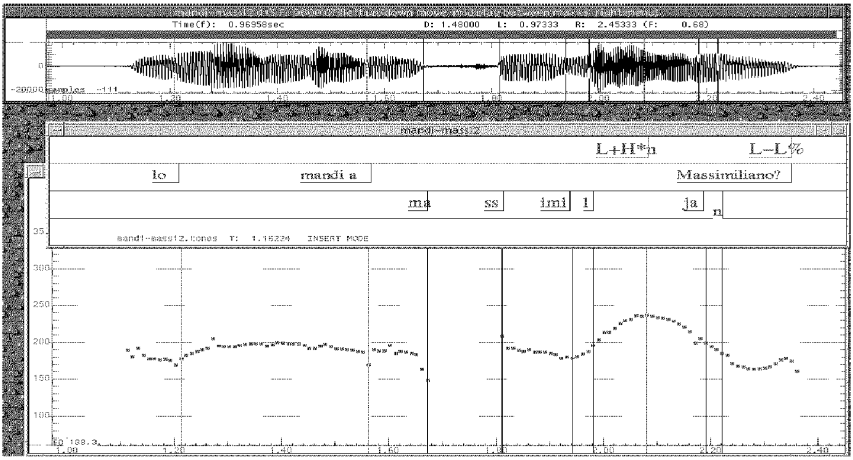


Figure 13.6 Bari: 'Lo mandi a [MassimiLIAno]?' (*Will you send it to [Maximillian]?*).

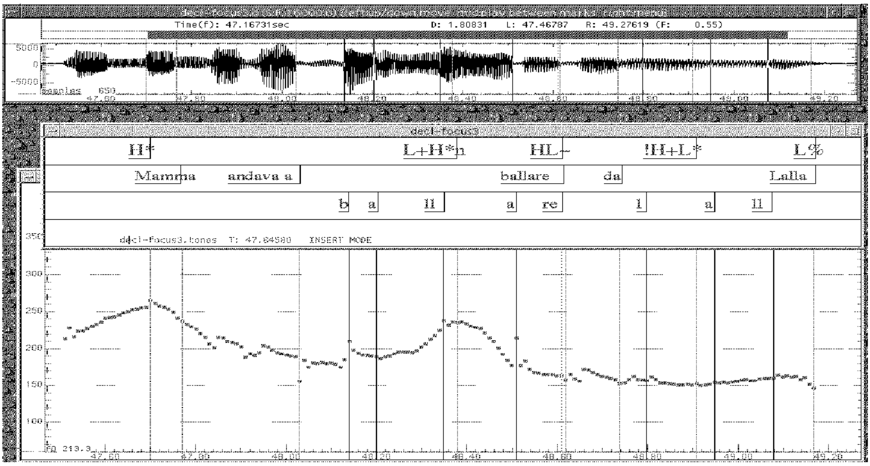


Figure 13.7 Neapolitan: 'MAMma andava a [balLAre] da Lalla' (*Mom used to go to[dance] at Lalla's*). Narrow focus declarative with L+H*.

cross-variety communication is hampered by the fact that the 'same' contour is used for different purposes. However, a closer look at the alignment details reveals that the L+H* accents are not identical in the two varieties. In fact, we can observe from Figure 13.7 and Figure 13.8 that although the valley and the elbow (the beginning of the steep rise) are in the same position in both varieties (respectively around the beginning of 'ballare' and at the onset of the

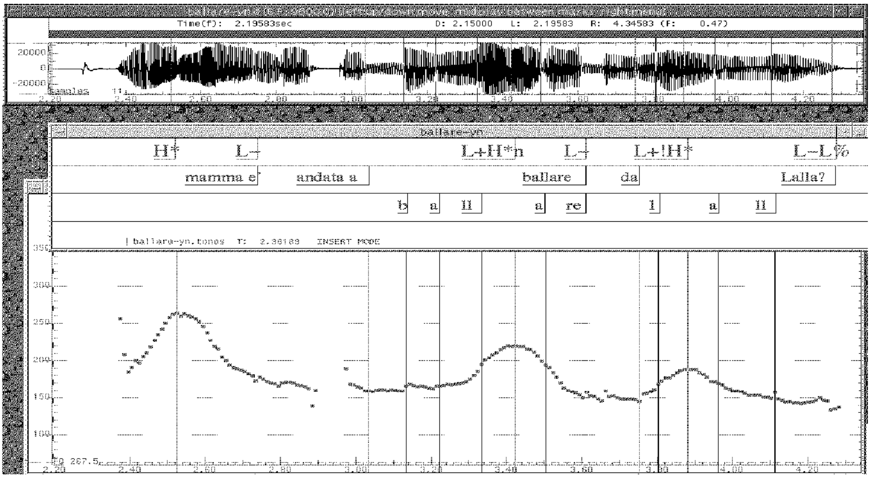


FIGURE 13.8 Bari: 'MAMma è andata a [balLAre] da LALla?' (*Did mom go to [dance] at Lalla's?*). Yes-no question with L+H*. Nuclear accent is on 'la' of 'ballare'.

stressed syllable), the peak is aligned differently: in Neapolitan it is early in the stressed vowel, whereas in Bari it is medial. This difference in peak timing leads to a perceived fall on the accented syllable in Neapolitan as opposed to a rise-fall in Bari. Also, while the fall of the Neapolitan L+H* is completely realized within the stressed vowel, this is not the case for Bari L+H*.

Recall that the Bari narrow focus pitch pattern has been transcribed as H*+L, especially when it is contrastive. This accent is also typical of confident checks, i.e. confirmation-seeking yes-no questions containing information which is confidently deemed by the speaker to be old (Grice and Savino 1997). An example of H*+L in a narrow focus declarative is given in Figure 13.9. A comparison between the Neapolitan narrow focus L+H* in Figure 13.7 above, and the Bari narrow focus H*+L in Figure 13.9 reveals that they have very similar contours.

Future research will have to decide whether a reanalysis of the Bari contour is necessary. A reanalysis as L+H* would of course mean that the question accent would have to have a different label. The choice of L*+H, despite the fact that the starred L* tone represents a valley occurring well before the accented syllable might be feasible if it were shown that the valley is consistently anchored to the accented syllable, despite its distance from it. That is, if the position of the valley is consistently related to the position of a landmark (e.g. vowel onset) in the accented syllable, it might be taken to be associated with it. A thorough investigation into the timing of the valley

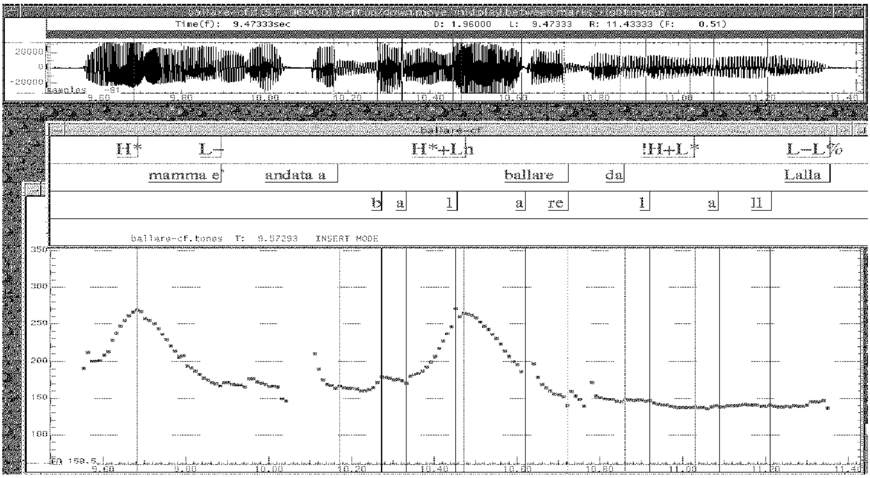


FIGURE 13.9 Bari: 'MAMma è andata a [balLAre] da LALla' (*Mom went to [dance] at Lalla's*). Contrastive narrow focus declarative with H*+L.

along the lines of Arvaniti *et al.* (1998, 2000) is clearly needed before this issue can be solved.

13.2.3. *Downstep*

Downstep is generally defined as a compression of the pitch range as a consequence of some phonological regularity. In the case of Southern varieties of Italian, we have observed that accents are downstepped in postfocal position, for both questions and statements, as for instance in !H+L* in Figure 13.2 and !H* in Figure 13.3 and Figure 13.4 shown above. Note that we did not include downstepped accents in our description of pitch accent types (see also Table 13.1) since we here follow Ladd (1996) in taking downstep to involve an orthogonal phonological variable independent of pitch accent type. Furthermore, although we take the downstepping of postfocal accents to be a predictable phenomenon, we nevertheless explicitly mark these accents as downstepped using the standard '!' symbol prefixed to the H tone of the accent concerned.

Also, we propose that, unlike in English, downstep applies across intermediate phrase boundaries (see Section 13.3 for a discussion of phrasing levels). That is, downstep is transcribed after the phrasal tone marking the edge of the focus constituent in Southern varieties, as for example in Figure 13.4 and

Figure 13.8 above. In this respect, therefore, Southern varieties are similar to Swedish, where downstep applies to postfocal accents, after the sentence accent (Bruce 1977).

13.3. LEVELS OF PHRASING

The varieties treated here are analysed as having two levels of phrasing relevant for intonational structure: the intonation phrase and a smaller phrase which is generally regarded to be akin to the intermediate phrase in English. Prosodic phonological analysis of Italian has delivered a degree of external evidence for the intonation phrase (Nespor and Vogel 1986; Frascarelli 1997) and in all of the varieties it is undisputed that it has a right peripheral tone which may be high (H%) or low (L%). In Florentine there is evidence for an optional left peripheral high tone (%H). By contrast, the intermediate phrase has only been tentatively proposed, based on the analysis of tonal configurations and a subjective impression of juncture.

13.3.1. *The intonation phrase*

All the varieties show a L% boundary tone at the right edge of the intonation phrase in declarative sentences (see the Neapolitan and Bari Italian examples in Figure 13.1 and Figure 13.2). A L% tone is also used to mark the right edge of the intonation phrase in yes/no questions in Neapolitan Italian (Figure 13.4). In Bari and Palermo Italian both H% and L% tones can be used to mark the right edge of such questions (see Table 13.2), while in Florentine only H% is attested (Figure 13.10).

In Florentine Italian there is also evidence of a left peripheral high boundary tone (%H), attested so far only in exclamative sentences. Various types of evidence support the claim that the high pitch start is a discrete, non-gradient event: the height of the first pitch accent does not affect the initial pitch height of the contour (see Figure 13.11, where three unstressed syllables precede the accent); resynthesized exclamative utterances with a lower initial pitch are no longer perceived as exclamatives; and, finally, the contour can be perceived independently of pitch range variations. It thus appears that the left peripheral %H boundary tone can be assigned a grammatical meaning, differently from Dutch, where its presence merely produces different pragmatic effects (Grabe *et al.* 1997). It is yet to be discovered whether the other varieties reveal such a distinction.

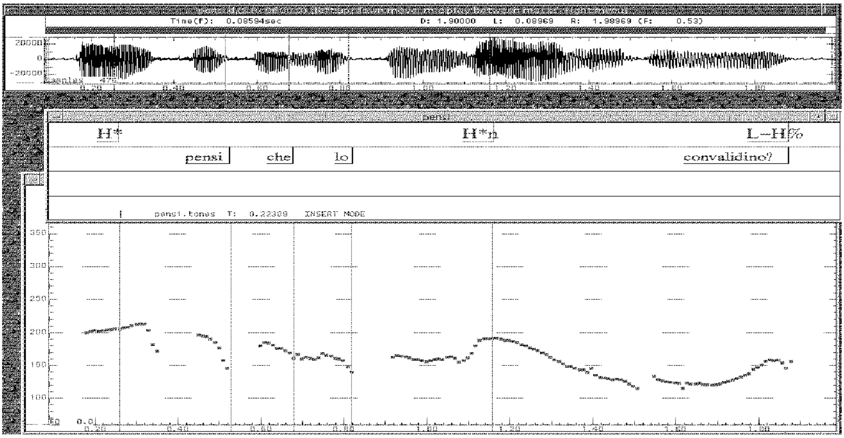


FIGURE 13.10 Florentine: 'Pensi che lo convalidino?' (*Do you think they will validate it?*). Yes/no question with a phrase final H% boundary tone.

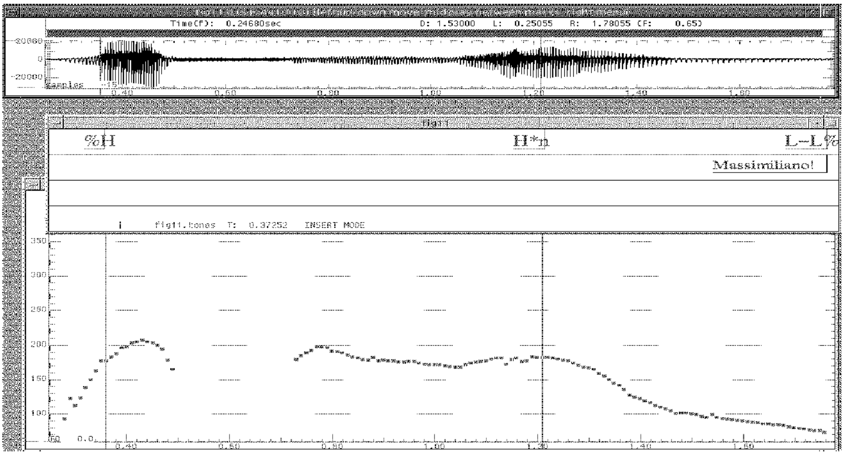


FIGURE 13.11 Florentine: 'MassimiLIAno!' (*Maximilian!*) Exclamative utterance with phrase-initial %H boundary tone.

13.3.2. *The intermediate phrase*

It is proposed for Florentine, Bari, and Palermo that the intermediate phrase is marked with a tone at its right edge (H- or L-). In the following section we examine the evidence for the intermediate phrase, taking as an example the Florentine variety. We then investigate how far the resulting analysis of Florentine can be extended to the other varieties.

(i) *The intermediate phrase in Florentine*: evidence for intermediate phrases comes first from the analysis of postposed vocatives (vocative tags) and right-disjoint adverbials (sententially-attached adverbials).

In both of these constructions the displaced element is perceived as prominent but is realized within a much lower range than the preceding item. In Florentine Italian, as in English (Beckman and Pierrehumbert 1986), the displaced element is treated as accented and is assigned a L* (Avesani 1995; Avesani and Hirschberg in prep.).

Tags are said to be prosodically in a closer relation with the main clause than a separate intonation phrase would be (Gussenhoven 1984; Beckman and Pierrehumbert 1986); on the other hand, it is also said that treating them as part of the same intonation phrase can be problematic. The same considerations can apply to the prosodic realization of Florentine tags and sententially attached adverbials: the latter are said to be syntactically adjoined to the sentence, but they are felt to be prosodically distinct from it, even if the sense of disjuncture between the adverbial and the sentence is less than what would induce the presence of an intonation phrase boundary. Positing the presence of an intermediate phrase boundary which separates the tag or the disjoined item from the rest of the utterance would account for these.

Consider for example a sentence like 'Accetta Mario' ('Accept Mario'). It can have two readings according to the function of the NP 'Mario': If 'Mario' is a direct object, it is nuclear in the intonation phrase [accetta Mario] and can be associated with a H+L* or H* accent depending on whether the sentence is meant to carry broad or contrastive focus. If 'Mario' is a vocative, its pitch contour displays only a minimal movement on the stressed syllable at a much lower range than the accent on the preceding VP (H*). Even if there are no pitch obtrusions (cf. Avesani 1995), 'Mario' is perceived as prominent, though it has a lower degree of prominence than the preceding VP.

Despite the absence of articulatory data which might provide evidence of the voluntary production of a pitch accent in a low range, it has been assumed that the prominence is due to the occurrence of a L* accent. The assumption is based on the following considerations:

(*a*) L* alternates with !H* in the same position. Even if less frequently, a postposed vocative or a sententially attached adverbial can have a clear pitch peak aligned with its stressed syllable. A postposed item with !H* is perceived as having a degree of prominence subordinated to the preceding accented item, just like postposed items lacking pitch obtrusions.

(*b*) In minimal pairs of VP-object/VP-vocative the duration of the stressed syllable in the vocative is not different from the duration of

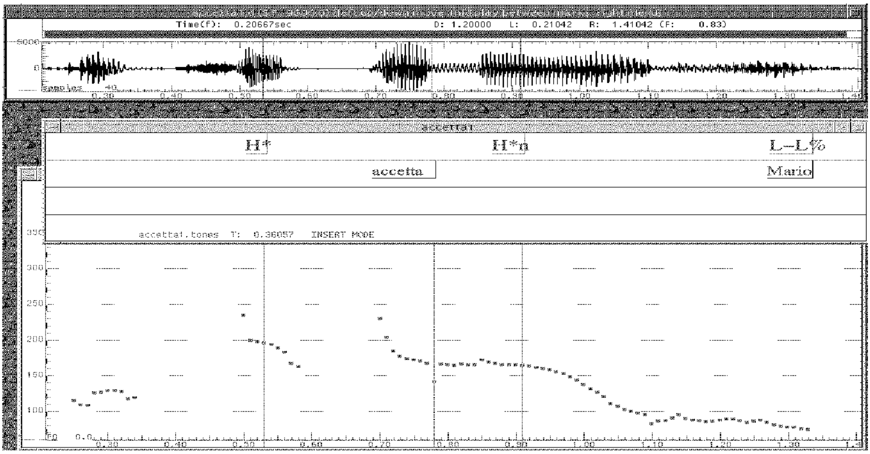


FIGURE 13.12 Florentine: 'Accetta Mario' (*Accept Mario*). Example of declarative sentence. The NP object 'Mario' is part of the same intermediate phrase as the VP 'accetta'.

> the stressed syllable of the accented object in the first member of the pair, which is associated with a H* or H+L*. The same holds true for minimal pairs of VP-attached/S attached adverbials (Hirschberg and Avesani 1997; Avesani 1999).

Figure 13.12 shows the contour for the object-reading of Mario in the sentence 'Accetta Mario', transcribed in (1): VP and NP objects are part of the same intermediate phrase. Figure 13.13 shows the contour for vocative-reading of 'Mario' in the same sentence, transcribed in (2): a L- tone sets apart the VP from the NP.

(1) acCETta MArio
H* H*n L-L%
'accept Mario' (Mario = object)

(2) acCETta MArio
H*n L- L* L-L%
'accept Mario' (Mario = vocative)

Sentences with left dislocation of the subject are another source of evidence. This type of dislocation can only be made possible if a weak prosodic break is inserted between the dislocated subject and the rest of the sentence. It is rhythmically realized with lengthening of the word final syllable—rarely via the insertion of a short pause—and melodically with a noticeable low or

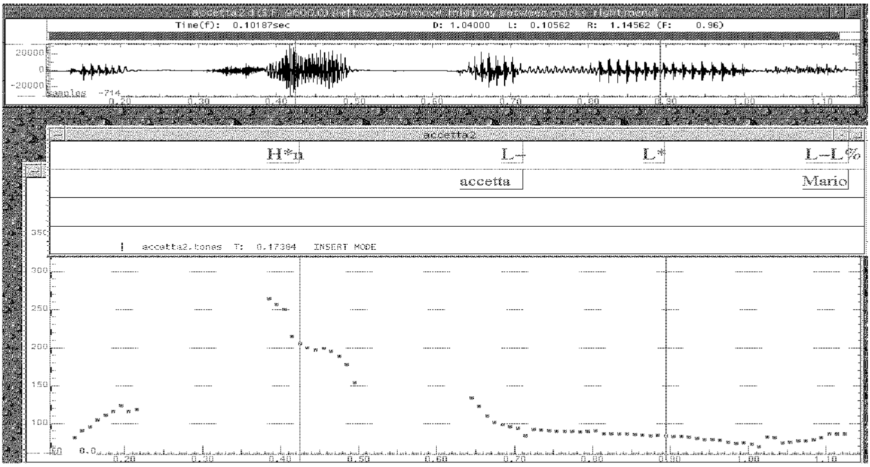


FIGURE 13.13 Florentine: 'Accetta Mario' (*Accept Mario*). The VP 'accetta' and the NP-vocative 'Mario' form different intermediate phrases.

rising pitch movement aligned with the last unstressed syllable (Avesani 1990). Left dislocation of the subject is more frequent with heavy NPs but perfectly possible and often attested in spontaneous speech also with light NPs.

Ambiguous syntactic attachment of prepositional and adverbial phrases and the ambiguous reading of relative clauses (restrictive vs. non-restrictive) are disambiguated by a different phrasing, mainly realized via the insertion of a H- or L- and only sometimes by a different duration of the lexical sequence across the critical syntactic boundary site (Avesani and Hirschberg in prep.).

Finally, the syntactic boundary of a conjoined clause may be marked by only a H- aligned with the last unstressed syllable of the clause (Avesani 1997).

(ii) *The intermediate phrase in other varieties*: in Bari Italian, evidence of an intermediate phrase boundary is found in yes-no questions read aloud with a following reporting clause such as example (3):

(3) 'Hai un eliporto?' ha chiesto allora Marcello
'"Do you have a heliport?" Marcello then asked'

The example is taken from a corpus of paragraph-length reported Map Task dialogues (see Savino and Refice 1996; Refice *et al.* 1997 for a description of the methodology used in eliciting such read materials). Two different strategies were found for reading out such a sequence. In the first, there was a strong sense of juncture after the question, which was pronounced with a final rise typical of questions read aloud (Grice *et al.* 1997; Refice *et al.* 1997). In the second strategy, there was a smaller but not negligible perceived juncture between the question and the reporting clause but the final rise was not realized until the end of the reporting clause. An indication that we are dealing here with a postponed question rise (Bolinger 1985) is given when comparing reporting clauses following questions with those following statements, as in example (4) (which follows (3) in the reading task):

(4) 'No, ho un aeroporto' ha risposto Giovanna
'"No, I have an airport" replied Giovanna'

where there is always a final fall instead of a rise at the end of the reporting clause. Furthermore, in cases where speakers produced a final rise at the end of the question, the reporting clause had a fall.

The second strategy is thus analysed as two intermediate phrases. Just as in isolated questions, the final rise is analysed as an intonation phrase H% tone, which must be placed at the intonation phrase boundary, which is at the end of the reporting clause, rather than at the end of the question where there is no available intonation phrase boundary. Thus:

Question alone	L+H* L-H%
Question+reporting clause	
Question	L+H* L-
Reporting clause	H+L* L-H%

Furthermore, intermediate phrase boundaries are transcribed for Bari and Palermo varieties after adverbials such as 'quindi', 'allora', and before the main clause in cleft sentences, as exemplified in (5):

(5) $[_{IP}[_{ip}$ È Giovanni $_{ip}]$ $[_{ip}$ che è partito? $_{ip}]$ $_{IP}]$
'Is it Giovanni who has left?'

The tonal analysis in Palermo Italian is shown in (6):

(6) È GioVANni che è parTIto?
L*+Hn L- L*+H L-L%

Since the Fo level reached at the end of the fall *Giovanni* is not as low as that reached at the end of *partito* (Grice 1995*a*), the boundary between the cleft and the main clause is taken to be a minor one. Following Pierrehumbert and Beckman (1988), the strength of the boundary is taken to affect the height of the associated tone.

13.3.3. *Break indices*

On the basis of the findings reported above for levels of phrasing, we propose five levels of break index: 0, 1, 2, 3 and 4. Break indices (BI) mark on a dedicated tier the perceived sense of disjuncture between words transcribed on the words tier. BI 0 should be used when two subsequent words show total cohesion, as in the case of a clitic group (Nespor and Vogel 1986), which is the prosodic constituent containing host and clitics (as, for instance [daˈlalla], from *da Lalla*, of Figures 13.2, 13.7, 13.8, and 13.9). BI 1 should be used to mark the disjuncture between clitic groups (such as the one between [mamma] and [andava] in Figure 13.7), while we reserve BI 2 (analogously to EToBI, Beckman and Ayers-Elam 1994) for cases where there is a mismatch between tonal and rhythmic cues to disjuncture. Finally BI 3 and BI 4 should be used to mark the discontinuity across intermediate phrases and intonation phrases respectively. Note that we do not include Break Indices in the figures shown. This is because the present paper concentrates specifically on tonal aspects of the varieties described.

13.4. WHAT IS A NUCLEAR PITCH ACCENT?

Autosegmental-metrical theory defines the nuclear accent as the last accent in a phrase. According to this definition, any lexical items following the nuclear accented word cannot bear an accent. Although Italian has a stronger tendency than English towards placing the nuclear accent late in the phrase (Grice 1995*a*; Ladd 1996), it does not rule out the placement of focus anywhere else in an utterance. This thus raises the question as to what happens to potential accents in postfocal words expressing given information. Unlike in English, where they would be deaccented (among others, Halliday 1967; Ladd 1980, 1996; Brown 1983; Cruttenden 1993), in Italian they are likely to be accented, since Italian tends to accent given information. For example, repeated lexical items are accented even if they share syntactic function and surface position with their antecedent expression (Avesani 1997). In fact, Ladd's (1996) claim that in Italian, as opposed to English, it is impossible to deaccent part of a syntactic phrase has been experimentally confirmed by Swerts *et al.* (1999). It appears that in postfocal position Italian only permits deaccenting of large syntactic constituents (full phrases or clauses). This means that the Italian focal accent can be optionally followed by other accents within the same phrase. This makes the broadly accepted positional definition of nuclear accent inadequate for Italian.

Instead, we take the Italian nuclear accent to be the rightmost fully-fledged pitch accent in the focused constituent. Since the focal structure is not necessarily labelled, and since other tones which bear resemblance to pitch accents may even occur within the focused constituent (see Section 13.5.1 below) we suggest appending a label 'n' to the nuclear pitch accent so that it is explicitly flagged. Any following pitch accent or tone within the same intonation phrase (whether in a separate intermediate phrase or not) is henceforth referred to as 'postnuclear' or 'postfocal'.

13.5. IS THERE A PHRASE ACCENT IN ITALIAN?

In this section we look at the evidence for analysing certain postnuclear tones in Neapolitan as boundary tones which have a secondary affiliation to a lexical stress, referred to as 'phrase accents'. We then consider how far this analysis can be taken to account for postnuclear accentual phenomena in other varieties.

13.5.1. *The phrase accent in Neapolitan*

A study of the focal accent in Neapolitan Italian (D'Imperio 1997, 2001) has shown some similarities with Swedish focal accent, which is marked by a separate tonal event, the phrase accent, originally referred to by Bruce (1977) as the 'sentence accent'. Since in both languages postfocal accents are not suppressed, the question arose as to whether a similar tonal event might exist in Italian. This question was addressed by investigating narrow focus patterns. When the constituent in focus is a single word, the fall of the interrogative rise-fall pattern occurs immediately after the pitch accent rise and appears to mark the end of the focus constituent ([*nano*] see Figure 13.3 above). When the focus constituent is longer, as in the case of ([*la bella mano di Mammola*] see Figure 13.4 above), the rise and fall appear to separate, with the rise staying anchored to the focus initial stressed syllable while the fall moves forward, reaching its target in the vicinity of the right-hand boundary of the constituent. From the above observation, it was hypothesized that the constituent final fall of interrogatives is analogous to the sentence accent of Swedish, in that this tone marks the end of the focus constituent and contributes to the perceived prominence of the focal accent, without creating the perceptual impression of a phrasal break. A production study concentrated on the properties of the final constituent fall in early focus interrogatives with different focus constituent sizes. It was found that the final HL fall is anchored to the last stressed syllable (when it is available) of multi-word focus constituents, thus resembling a regular pitch accent. It is important to stress here that the final fall is taken to only resemble a pitch accent. If it were an ordinary pitch accent, then even the new definition for the nucleus as final pitch accent in the focused constituent would be invalid, since the nuclear pitch accent here is L*+H.

When there is only one stressed syllable in the focus constituent, the nuclear L*+H will take over, leaving the HL sequence to be realized as an appendix of the rise. Specifically, it was found that the target for the HL fall is reached later in single-word focus constituents, as if it were 'pushed' outside the stressed syllable by the nuclear L*+H. This is another indication for the L*+H having a primary and therefore in this case nuclear association to the stressed syllable of the focused word.

13.5.2. *Can the phrase accent analysis be extended to the other varieties?*

The question arises as to whether the postfocal tones found in the other varieties can be analysed as phrase accents. There are two types of postfocal tone: those where the nucleus and the postfocal tone are within the same phrase, and those where an intermediate phrase boundary intervenes. The first type is found in Palermo Italian. Grice (1995*a*) discusses a contour in Palermo Italian used in confirmation-seeking yes-no questions, which is reproduced with its original transcription in (7).

(7) TU gliel'hai DETto?
L*+H H+L* L-L%
'You said it to him/her?'

Here the nuclear accent is L*+H. The contour could at first glance be reanalysed as L*+H HL-L%, as in Neapolitan. However, the association of the second H tone is different. In Neapolitan, the bitonal HL- phrase accent aligns the (H) shoulder with the stressed syllable, whereas in Palermo it is the (L) valley which is aligned with the stress. However, the HL- analysis is appropriate if the alignment is captured by a strength relation between the two components of a branching ip edge tone:[4]

(8) (*a*) ip
/\
H_s L_W in Neapolitan

[4] Note that a similar branching phrase accent with a weak-strong relation between the tones has been independently proposed for a dialect of German (Peters 2001).

(*b*) ip
/\
H_w L_s in Palermo

This analysis entailing a branching edge tone is analogous to the analysis of branching pitch accents (Pierrehumbert and Beckman 1988; Grice 1995*b*) where the starred tone of a bitonal pitch accent is represented as strong, as in the representation of L*+H in (9).

(9) pitch accent
/\
L_s H_w

In the proposed annotation scheme, 8(*a*) would be transcribed as H(*)L-, and 8(*b*) as HL(*)-. We capture instances of secondary association with a stressed syllable by means of a parenthesized star (*), and the primary intermediate phrase association by retaining the original '-' symbol (which is placed after the second tone if the boundary is bitonal). The parenthesized star is currently only present when there is a stressed syllable with which the phrase accent is associated. The transcription of (7) with a phrase accent rather than a postfocal pitch accent is given in (10).

(10) TU glielʼhai DETto?
L*+Hn HL(*)- L%

Analogously, observe that the phrase accent in Neapolitan is transcribed in Figure 13.4 as H(*)L-. Note that in the absence of a syllable for secondary association, it is transcribed HL-, as in Figure 13.5.[5]

The other type of postfocal tone occurs after a phrase boundary. For instance, in the Florentine example in (2) above, reproduced here as (11), the second transcribed accent could be reanalysed as a phrase accent L- which associates with the stressed syllable 'MA' of 'MArio' instead of the pitch accent L* in the original analysis.

(11) acCETta MArio
H*n L- L* L-L% (original analysis)
H*n L- L(*)- L% (alternative analysis)

Here too, the proposed phrase accent is different from the Neapolitan version, this time because it is monotonal and because an intermediate phrase

[5] The postnuclear !H+L* accents in the Bari examples given in Figure 13.2 and Figure 13.9 could also be analysed as HL(*)- phrase accents. We leave this question open for further research, especially given that the analysis of the type of contrastive focus utterance presented in Figure 13.9 is still tentative.

boundary intrudes between the nuclear accent and the phrase accent. This new analysis is inspired by Gussenhoven's (1990) tone copy analysis of intonational tags, such as in reporting clauses, where the unstarred tones of the melody in the main clause are copied to the tag.

```
(12)        Were you THERE?    asked JONathon
                     H*LH             L      H
                                   adapted from (Gussenhoven 1990: 35)
```

There is, however, another type of postfocal tone where an analysis in terms of phrase accents is not an obvious option, namely in yes-no questions in Bari and Palermo. In each variety the postfocal accent is very similar in shape and timing to the focal accent it follows (which, recall, is L+H* in Bari and L*+H in Palermo), and is taken to be a copy of it in a reduced pitch range. These postfocal tones are analysed as downstepped pitch accents (L+!H* and L*+!H respectively). An example of post-focal L+!H* in Bari Italian is shown in Figure 13.8 on the word 'LALla' in the narrow focus yes-no question 'MAMma è andata a [balLAre] da LALla?'. Such a post-focal accent plays a basic role in terms of meaning: if it is suppressed (by using resynthesis, for example), the utterance is no longer perceived as a question.

As we have seen, not all postfocal tones which have an association to a stressed syllable can be reanalysed as phrase accents. In cases where a phrase accent analysis is presumed, we have suggested transcribing the associated tone with a parenthesized star. The affiliation of the tone with a phrase boundary (by means of a '-' symbol) is left untranscribed only in those varieties in which the distribution of these tones is still unclear. The 'n' appended to L*+H, marking the identity of the nuclear pitch accent, is particularly important in the labelling scheme, since the nuclear pitch accent is neither the last apparent pitch accent in the focused constituent (e.g. H(*)L- in Neapolitan in Figure 13.4), nor the last accent in the phrase (e.g. !H* in Figure 13.3 and !H+L* in Figure 13.9).

13.6. TRUNCATION IN THE SOUTHERN VARIETIES

It is worth noting that in the Southern varieties, phrasal tones are not always fully realized if the final syllable of a phrase is accented, as in the Bari example in Figure 13.14, where the final syllable of *Noè* is associated with a L+H* pitch accent. The pitch accent is followed by two edge tones, L- and L%, which must be realized on the same syllable. A full realization of these tones would mean a fall to low in the range, as is usually the case in yes-no questions in

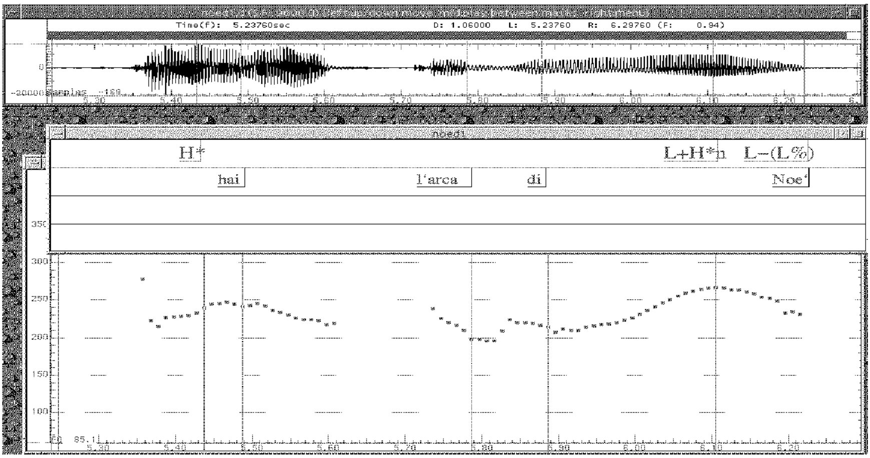


FIGURE 13.14 'Hai l'arca di noÈ?' (*Do you have Noah's ark?*). Truncated rise-fall in phrase-final accented syllable in Bari Italian.

this speaking style (Grice *et al.* 1997; Savino 1997). Instead of a full realization, the fall only reaches a level around the middle of the range, and is analysed as being truncated (see Grice 1995*a* for Palermo Italian; Grice *et al.* 1997; Refice *et al.* 1997 for Bari Italian). Our proposal for the ToBI labelling of Italian varieties is to explicitly flag truncation by placing partially realized tones in round brackets. An example of this notation is shown in Figure 13.14, where the fall to mid is captured by (L%).

13.7. CONCLUSION

Despite the lack of a widely recognized Standard, we have shown that it is possible to find common traits among the intonation structure of various regional varieties of Italian, three from the South (Neapolitan, Bari, and Palermo) and a Central one (Florentine). First, we have shown that these varieties share a basic set of intonational elements and appear to organize them in similar ways. For instance, in all of the varieties examined, nuclear pitch accent type is used to distinguish contrastive narrow focus from broad focus in declaratives. This distinction is captured in all of the systems reported upon, although, despite an apparent phonetic similarity across the varieties, there are three different ways of transcribing the narrow focus accent. We suggest ways of clarifying whether or not these differences are justified by the phonetic facts.

Additionally, all of the Southern varieties examined employ a rising L+H pitch accent in yes-no questions as principal indicator of interrogativity. On the grounds of functional identity, we discuss whether all three varieties should be transcribed with the same accent type even though there are obvious differences across the varieties in the alignment of the H peak. We opt for reflecting the alignment in the transcription, at least until further evidence of contrasting accents within a variety lead us to reanalyse the starredness of the tones.

Regarding phrasing levels, we have presented evidence supporting the existence of an intermediate phrase and an intonation phrase level, though more data is needed to support the intermediate phrase level. On the basis of the findings, we propose the use of five Break Indices (0, 1, 2, 3, and 4), directly reflecting the phrasing levels discussed here as well as a prosodic level that has been widely studied within the Prosodic Phonology framework (Nespor and Vogel 1986). We assume that such phrasing levels are common to all the varieties of Italian presented here.

We have also discussed the issue of postnuclear accents, which have been attested in some form in all of the varieties, and have proposed flagging the nuclear accent with an appended 'n' so that nuclear and postnuclear accents can be easily and unambiguously identified. The advantage of keeping the labelling scheme surface oriented and theoretically conservative is that more research into the underlying phonological structure can be undertaken without updating label files each time new results are obtained. It also allows for the labelling of a database without opting necessarily for one of a number of competing phonological analyses, thus making it more versatile and more broadly usable.

Finally, we have proposed that a process of downstep applies in Southern varieties, modifying the range of pitch accents in postfocal position. Postfocal accents occur after the nuclear accent, either within the same phrase or across phrases. In both situations they appear to be compressed in a way that makes them almost indiscernible. Further research is needed in order to determine whether downstep applies in other conditions and environments too.

REFERENCES

ARVANITI, A., and BALTAZANI, M. (this volume Ch. 4), 'Intonational Analysis and Prosodic Annotation of Greek Spoken Language Corpora'.

— —, LADD, D. R., and MENNEN, 1. (1998), 'Stability of Tonal Alignment: the Case of Greek Prenuclear Accents', *Journal of Phonetics*, 26: 3-25.

ARVANITI, A., LADD, D. R., and MENNEN, 1. (2000), 'What is a Starred Tone? Evidence from Greek', in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (Cambridge: Cambridge University Press), 119-3l.

AVESANI, C. (1990), 'A Contribution to the Synthesis of Italian Intonation', in Congress Committee and Organising Committee (eds.) *Proceedings ofICSLP 90, International Conference on Spoken Language Processing* (Kobe), 834-6.

— — (1995), 'ToBIt. Un sistema di trascrizione per l'intonazione italiana', in G. Lazzari (ed.), *Atti delle V Giornate di Studio del Gruppo di Fonetica Sperimentale dell' Associazine Italian di Acustica* (Trento: Servizio Editoria lTC), 85-98.

— — (1997), 'I toni della RA1. Un esercizio di lettura intonativa', in G. Nencioni and N. Maraschio (eds.), *Gli italiani trasmessi: la radio* (Florence: Accademia della Crusca), 659-727.

— — (1999), 'Quantificatori, negazione e costituenza sintattica. Costruzioni potenzialmente ambigue e il ruolo della prosodia', in P. Benincà, A. Mioni, and 1. Vanelli (eds.), *Atti del XXXI Congresso Internazionale di Studi della Societa di Linguistica Italiana* (Rome: Bulzoni), 153-200.

— —, and HIRSCHBERG, J. (in prep.), 'The Prosodic Disambiguation of Syntactically and Semantically Ambiguous Sentences in English and Italian'.

— —, and VAYRA, M. (2000), 'Strutture marcate e non marcate in italiano. II ruolo dell'intonazione', in D. Locchi (ed.), *Atti delle* X *Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italian di Acustica* (Naples: Istituto Universitario Orientale), 1-14.

BECKMAN, M., and AYERS-ELAM G. (1994), 'Guidelines for ToBI labelling, version 3.0, March 1997', ms, Ohio State University.

— —, and PIERREHUMBERT, J. (1986), 'Intonational Structure in English and Japanese', *Phonology Yearbook*, 3: 255-310.

BERTINETTO, P. M. (1981), *Strutture prosodiche dell'italiano* (Florence: Accademia della Crusca).

BOLINGER, D. (1985), *Intonation and its Parts* (Ann Arbor: E. Arnold).

BROWN, G. (1983), 'Prosodic Structure and the Given/New Distinction', in A. Cutler and R. Ladd (eds.) *Prosody: Models and Measurements* (Berlin: Springer-Verlag), 67-78.

BRUCE, G. (1977), *Swedish Word Accents in Sentence Perspective* (Lund: Gleerups).

CARLETTA, J., ISARD, A., ISARD, S., KOWTKO, J. C., DOHERTY-SNEDDON, G., and ANDERSON, A. H. (1997), 'The Reliability of a Dialogue Structure Coding Scheme', *Computational Linguistics,* 23/1: 13-3l.

CRUTTENDEN, A. (1993), 'The De-Accenting and Re-Accenting of Repeated Lexical Items, in D. House and P. Touati (eds.), in *Proceedings of ESCA Workshop on Prosody* (Lund: Reprocentralen Lund University), 16-19.

D'IMPERIO, M. (1997), 'Narrow Focus and Focal Accent in the Neapolitan Variety of Italian', in A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.), *Intonation:*

Theory, Models and Application. Proceedings of ESCA Workshop on Intonation (Athens: Athanasopoulos and Papadamis), 87-90.

D'IMPERIO, M. (1999), 'Tonal Structure and Pitch Targets in Italian Focus Constituents', in J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds.), *Proceedings of the XIV International Congress of Phonetic Sciences* (Berkeley: University of California), 3: 1757-60.

— — (2001), 'Focus and Tonal Structure in Neapolitan Italian', *Speech Communication*, 33(4): 339-56.

— —, and ROSENTHALL S. (1999), 'Phonetics and Phonology of Main Stress in Italian', *Phonology*, 16/1: 1-28.

ENDO, R., and BERTINETTO, P. M. (1997), 'Aspetti dell'intonazione in alcune varieta dell'italiano', in F. Cutugno (ed.), *Atti delle VII Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italiana di Acustica* (Rome: Esagrafica), 27-49·

FARNETANI, E., and BUSÀ, M. G. (1999), 'Quantifying the Range of Vowel Reduction in Italian', in J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds.), *Proceedings of the XIV International Congress of Phonetic Sciences* (Berkeley: University of California), 1: 491-4.

— —, and KORI, S. (1986), 'Effects of Syllable and Word Structure on Segmental Durations in Spoken Italian', *Speech Communication*, 5: 17-34.

FRASCARELLI, M. (1997), 'The Phonology of Focus and Topic in Italian', *The Linguistic Review*, 14: 221-48.

GALLI DE' PARATESI, N. (1985), *Lingua toscana in bocca ambrosiana. Tendenze verso l'italiano standard: un inchiesta sociolinguistica* (Bologna: II Mulino).

GILl FIVELA, B. (2003), 'The Phonetics and Phonology of Intonation: the Case of Pisa Italian', Ph.D. dissertation (Scuola Normale Superiore, Pisa).

GRABE, E., GUSSENHOVEN, C., HAAN, J., MARSI, E., and POST, B. (1997), 'Preaccentual Pitch and Speaker Attitude in Dutch', *Language and Speech*, 41/1: 63-85.

GRICE, M. *(1995a), The Intonation of Palermo Italian; Implications for Intonation Theory* (Tiibingen: Niemeyer Linguistische Arbeiten 334).

— — *(1995b)*, 'Leading Tones and Downstep in English', *Phonology*, 12: 183-233.

— —, and SAVINO, M. (1997), 'Can Pitch Accent Type Convey Information Status in Yes-No Questions?', in K. Alter, H. Pirker, and W. Finkler (eds.), *Proceedings ofthe ACL Workshop 'Concept-to-Speech Generation Systems'* (Madrid: UNED), 29-38.

— —, — —, and REFICE, M. (1997), 'The Intonation of Questions in Bari Italian: Do Speakers Replicate Their Spontaneous Speech when Reading?', *Phonus*, 3: 1-7.

GUSSENHOVEN, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

— — (1990), 'Tonal Association Domains and the Prosodic Hierarchy in English', in S. Ramsaran (ed.), *Studies in the Pronunciation ofEnglish* (London: Routledge), 27-37·

HALLIDAY, M. A. K. (1967), 'Notes on Transitivity and Theme in English. Part 2', *Journal ofLinguistics*, 3: 199-244·

HIRSCHBERG, J., and AVESANI, C. (1997), 'The Role of Prosody in Disambiguating Potentially Ambiguous Utterances in English and Italian', in A. Botinis, G. Kouroupetroglou, and G. Karayiannis (eds.), *Intonation: Theory, Models and Applications, Proceedings of ESCA Workshop on Intonation* (Athens: Athanasopoulos and Papadamis), 189-92.

LADD, D. R. (1980), *The Structure of Intonational Meaning: Evidence from English* (Bloomington: Indiana University Press).

— — (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

LEPSCHY, A. 1., and LEPSCHY, G. C. (1977), *The Italian Language Today* (London: Hutchinson).

MAROTTA, G. (2000), 'Allineamento e trascrizione dei toni accentuali complessi: una proposta', in D. Locchi (ed.), *Atti delle* X *Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italiana di Acustica* (Naples: Istituto Universitario Orientale), 139-49.

— —, and SORIANELLO, P. (1999), 'Question Intonation in Sienese Italian', in J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds.), *Proceedings of the XIV International Congress ofPhonetic Sciences* (Berkeley: University of California), 2: 1161-4.

NEsPoR, M., and VOGEL, 1. (1986), *Prosodic Phonology* (Dordrecht: Foris).

PETERS, J. (2001), 'Postnukleare Tonhohengipfel in der Vorderpfalz und in Mannheim', ms, University of Potsdam.

PIERREHUMBERT, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology) (Bloomington: Indiana University Linguistics Club (1987)).

— —, and BECKMAN, M. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

RAMUS, F., NEsPoR, M., and MEHLER, J. (1999), 'Correlates of Linguistic Rhythm in the Speech Signal', *Cognition,* 73: 265-92.

REFICE, M., SAVINO, M., and GRICE, M. (1997), 'A Contribution to the Estimation of Naturalness in the Intonation of Italian Spontaneous Speech,' in G. Kokkinakis, N. Fakotakis, and E. Dermatas (eds.), *Proceeding of the V European Conference on Speech Communication and Technology* (Rhodes: WCL, University of Patras), 2: 783-6.

ROMITO, 1., and TRUMPER, J. (1989), 'Un problema della coarticolazione: l'isocronia rivisitata', in Congress Committee and Organising Committee (eds.), *Atti del XVI Convegno dell'Associazione Italiana di Acustica* (Fidenza: Tipolitografia Mattioli), 449-55·

SAVINO, M. (1997), 'II ruolo dell'intonazione nell'interazione comunicativa. Analisi strumentale delle domande polari in un corpus di dialoghi spontanei (varieta di Bari)', Ph.D. dissertation (Università/Politecnico di Bari).

— —, and REFICE, M. (1997), 'L'intonazione dell'italiano di Bari nel parlato letto e in quello spontaneo', in F. Cutugno (ed.), *Atti delle VII Giornate di Studio del Gruppo di Fonetica Sperimentale dell'Associazione Italiana di Acustica* (Rome: Esagrafica), 79-88.

SWERTS, M., AVESANI, C., and KRAHMER, E. (1999), 'Reaccentuation or Deaccentuation: a Comparative Study of Italian and Dutch', in J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Bailey (eds.), *Proceedings of the XIV International Congress of Phonetic Sciences* (Berkeley: University of California), 2: 1541-4.

VAYRA, M., AVESANI, C., and FOWLER, C. (1984), 'Patterns of Temporal Compression in Spoken Italian', in M. van den Broecke and A. Cohen (eds.), *Proceedings of the* X *International Congress of Phonetic Sciences* (Dordrecht: Foris), 2: 540-6.

— —, FOWLER, C., and AVESANI, C. (1987), 'Word-Level Coarticulation and Shortening in Italian and English Speech', in Haskins Laboratories Status Report on Speech Research SR/91, 75-89. Also in *Studi di Grammatica Italiana,* 13: 249-69.

14

Intonational Variation in Four Dialects of English: the High Rising Tune

Janet Fletcher, Esther Grabe, and Paul Warren

14.1. INTRODUCTION

Many varieties of English have intonational patterns or patterns of 'tune usage' that differ from the most studied varieties, namely General American English and Standard Southern British (SSB) English (Cruttenden 1994; Ladd 1996; Grabe *1998a, b,* Grabe *et ai.* 1998). In this chapter we propose to show how intonational analyses within an autosegmental-metrical phonological framework can deal with this type of variation. In Sections 14.2.1 and 14.2.2 of this chapter we re-examine the nuclear rising tunes often associated with declarative utterances in Australian English and New Zealand English with particular focus on socio-phonetic differences. Section 14.3 describes the nuclear rises in Glasgow English and Belfast English compared to other varieties of British English. As pointed out by Cruttenden (1994), these rises differ from the declarative rises of the two Australasian varieties in that they constitute a 'standard' declarative tune rather than what has traditionally been considered to be a sociophonetically marked pattern in New Zealand and Australia. They also differ somewhat in their phonetic realization.

We propose to re-examine intonational differences among varieties of English in terms of a typology elaborated by Ladd (1996) which is based partly on a descriptive framework established by Wells (1982) to describe segmental differences among varieties of a single language. Ladd's typology was developed primarily to consider intonational differences among languages but he also applies it to varieties within a language. We propose to extend his treatment of dialectal differences to consider issues of sociophonetic variation within a variety, and to consider the possibility of differences also being neutralized in certain situations. The Ladd typology includes semantic differences (Type 1) which cover differences in meaning or function of phonologically identical tunes; systemic differences (Type 2) which include differences in the inventory of phonologically distinct tune types, irrespective of Type 1 differences, realizational differences (Type 3) where phonetic realization of an identical phonological tune may differ among or within varieties, and finally phonotactic differences (Type 4) which are differences in tune-text association.

In the final section of this chapter we consider representational issues that emerge from analysis of cross-dialectal phenomena and we include brief descriptions of two annotation systems that have emerged in recent years.

14.2. INTONATION IN AUSTRALIAN ENGLISH AND NEW ZEALAND ENGLISH

14.2.1. *The high rising terminal: a semantic difference?*

Earlier studies of Australian English (e.g. Mitchell and Delbridge 1965) have proposed that Australian speakers use a similar inventory of intonational patterns to speakers of Standard Southern British (SSB) English or standard American English. This suggests that the intonational phonology of Australian English is not significantly different from SSB English. However, potential cross-variety differences need to be taken into consideration with respect to tune usage. Previous studies of Australian English (e.g. Guy and Vonwiller 1984, 1989; Horvath 1985) document the growing usage since the 1970S of the so-called high rising terminal (henceforth referred to as the HRT) accompanying syntactically marked declarative utterances as well as yes/no questions. A similar phenomenon has been widely documented for its near dialectal neighbour, New Zealand English (e.g. Britain 1992; Cruttenden 1994). It is widely claimed that phonetically identical high-rising tunes can be used to signify these two different utterance types, which makes Australian English and New Zealand English intonation different from SSB English, for example where the high-rising nucleus is used primarily with yes-no questions, and never with declarative utterances. Ladd (1996) classifies this dialectal difference as an example of a 'semantic' or Type 1 difference.

The majority of research on the use of HRT with declarative clauses in Australian English has been sociophonetic and auditory impressionistic. Halliday's 'phonetic' definition of an HRT is that the tune must rise to a pitch level approximately 40 per cent higher than the high nuclear accent. In

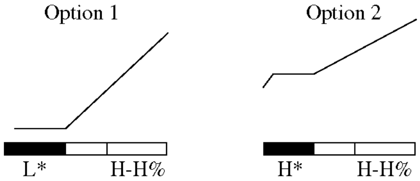


FIGURE 14.1 Illustration of two kinds of rising tunes in Australian English.

autosegmental-metrical terms (after Pierrehumbert 1980), the transcription of this tune would be H* H-H%, although Ladd suggests that the L* H-H% tune might also be classed as an HRT. Figure 14.1 illustrates these two options for Australian English. The first tune is associated with an observable Fo trough in the low part of a speaker's range associated with the primary stressed syllable of the accented word, rising to relatively high pitch level at the edge of the intonational phrase. By contrast, the second tune commences relatively high in the speaker's range, and continues to rise towards the phrase edge.

For the most part, researchers of the phenomenon in Australian English have adhered to Halliday's original definition of an HRT as a high rising nuclear tone, following transcription conventions of the British School of Intonation (e.g. Halliday 1967). Some consensus emerges from the sociolinguistic studies of the 1980s that the tune is used predominantly by young adolescent females, is associated with low prestige varieties (i.e. 'broad' Australian English), is socially stigmatized, occurs most often in the telling of narratives, and is spreading through other sections of the Australian community. Guy and Vonwiller (1989: 25) suggest 'HRT correlates with the semantic complexity of the text and therefore the need for checking to see if the audience is understanding what is being said'. Speakers tend to use it as a device to hold, rather than yield the floor in discourse situations. Guy and Vonwiller also claim that HRT usage fits with more general interpretations of rising pitch as signifying non-finality.

The situation in New Zealand English appears to be somewhat similar. As in Australian English, HRTs are more typically associated with narratives than with opinion texts, and in New Zealand English they are also more frequent amongst women and speakers of Maori ethnicity (Britain 1992). There is a suggestion that HRTs also serve as positive politeness markers in New Zealand English, maintaining speaker-hearer solidarity (for a review see Warren and Britain 2000).

For Australian English, anecdotal evidence suggests that the tune is no longer primarily associated with young adolescent female Australians, and has been adopted by a broader cross-section of the Australian community, as predicted by the sociolinguistic studies of the 1980s. Data from the ANDOSL (Australian National Database of Spoken Language) MAP corpus (Millar *et al.* 1994) will be presented to show that male and female participants in the MAP task frequently use the HRT. We will also show that HRTs labelled as either L* H-H% or H* H-H% are used with declarative utterances as well as with yes/no questions. The ANDOSL MAP task[1] is similar to the original HCRC map task developed in Edinburgh and Glasgow in the late 1980s (Anderson *et al.* 1991) and has been used widely in intonational studies of other languages, like German and Japanese. Versions of a map task have also been developed for other varieties of English including Northern British varieties (Grabe *et al.* 2000) and New Zealand English (Daly and Warren 2001). In the ANDOSL corpus, male and female speakers have been recorded from the three linguistically defined dialectal groups of Australian English—cultivated, general, and broad (Mitchell and Delbridge 1965). The following examples of HRTs have been chosen from a section of the ANDOSL MAP database that has been fully transcribed according to autosegmental-metrical conventions that are more or less equivalent to the Tones and Break Indices (ToBI) tagging criteria outlined by Pitrelli *et al.* (1994) for American English. The conventions for Australian English are outlined in Fletcher and Harrington (2001). All data presented below are from speakers of 'general' Australian English.

Figures 14.2(a) and (b) show two examples of tunes labelled H* H-H% from a female speaker. The first contour shows the rising tune with a yes/no question and the second contour shows a declarative utterance with essentially the same tune. Similarly Figures 14.3(a) and (b) show a male speaker illustrating two instances of the L* H-H% 'low-rising' tune, one for a declarative utterance and the other for a yes/no question. The tunes are identical, apart from a slightly higher Fo value at the endpoint of the contour associated with the yes/no question.

Intonational phrases labelled L* H-H%, L+H* H-H% or H* H- H% were all classified as examples of HRTs, following a suggestion made by Ladd (1996). Looking at a sample of the labelled corpus (three females, six males), 21 per cent of all high rising tunes produced by the female speakers coincide with declarative utterances, whereas for males, the proportion is slightly higher at 25 per cent. The male speakers use HRTs more often than females, contrary to earlier sociolinguistic findings (e.g. Horvath 1985; Guy and

[1] Participants in the ANDOSL MAP task work in pairs, each with a map in front of them that the other cannot see. One participant (the 'instruction-giver' IG) has a route marked on their map and is required by the task to instruct the other (the 'instruction-follower' IF) in drawing the correct route onto their own map. The maps are similar, but differ in the presence, position, and names of certain of the landmarks.

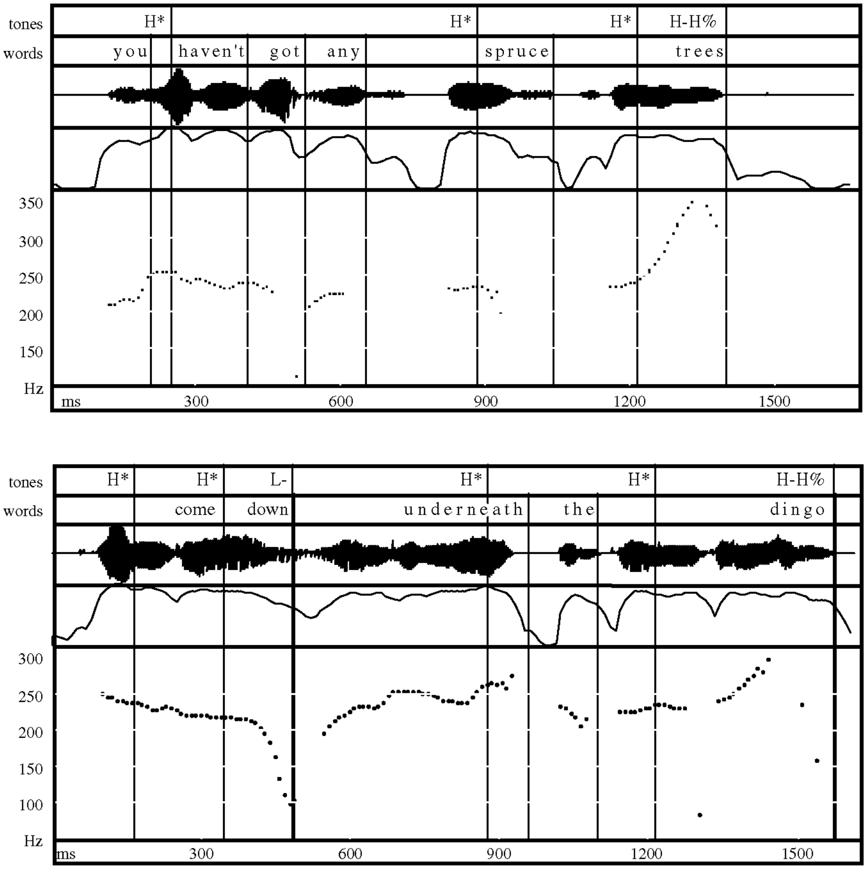


FIGURE 14.2 (a) An example of an H* H-H% high rising tune from Australian English. This example is produced by a female speaker and accompanies the yes/no question 'You haven't got any spruce trees?' (b) An H* H-H% nucleus produced by a female performing the 'leader' role in a map task dialogue. The utterance is part of an instruction to the 'follower' in the map task, 'comes down underneath the dingo'.

Vonwiller 1989). The rest of the H-H% tunes are generally associated with yes/no questions. It is also interesting to note that these results are from speakers performing the 'leader' role in the MAP task. Generally, more H-H% tunes associated with yes/no questions are observed when speakers are adopting the role of the 'follower'. It is also not surprising that HRTs are found in this kind of task. Its construction is specifically designed so that one participant is frequently requesting information, or seeking verification and confirmation from the other participant. A MAP task can also be construed as a 'semantically'

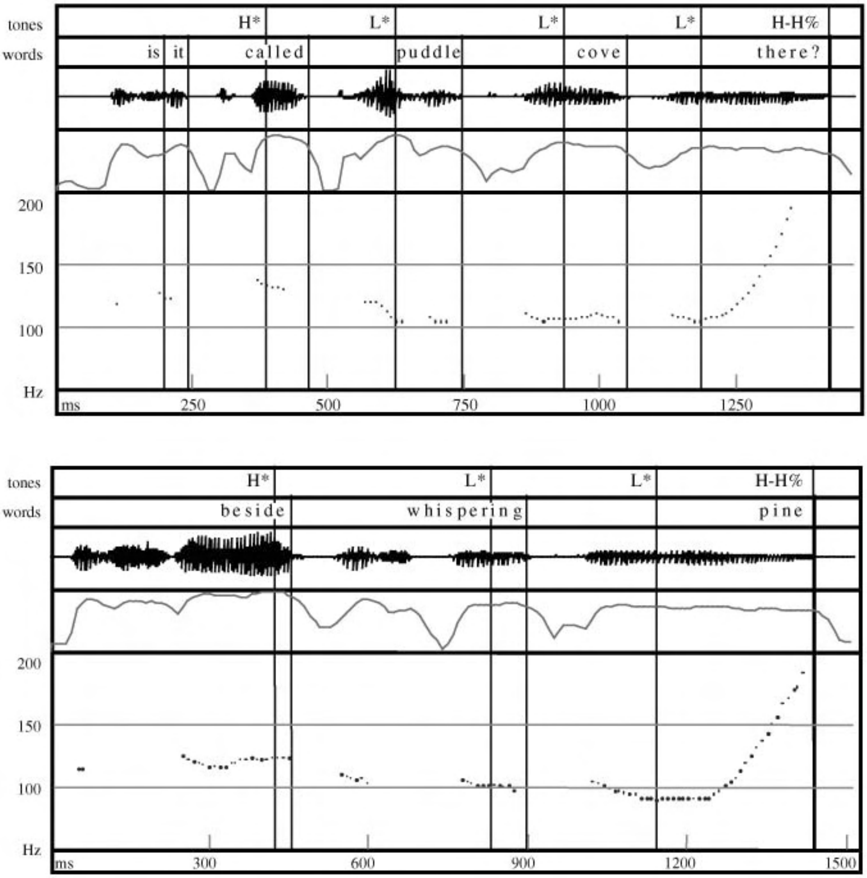


FIGURE 14.3 (a) An example of a L* H-H% 'low' rising tune produced by a male speaker of Australian English. This example is a question, 'Is it called puddle cove there?' (b) This example of a L* H-H% nucleus is part of a statement: '(it should be) beside whispering pine'. Note the similar extent of the rise from low to high pitch across this and the previous example.

complex text or a type of narrative. The results reported here therefore concur with earlier findings (e.g. Horvath 1985; Guy and Vonwiller 1989) for Australian English, namely, narratives are associated with a higher incidence of HRTs than non-narrative or 'opinion' texts.

There is an interesting trend in some of the MAP data analysed so far. It is apparent that there are no discernible phonetic differences between the HRTs associated with questions and declaratives for speakers like the one whose rises are illustrated in Figures 14.3(a) and (b), supporting earlier claims (e.g. Ladd

1996: 121). However other speakers in the MAP corpus systematically use 'high' nuclear rises for questions (i.e. H* H-H%) and 'low' nuclear rises for statements (L* H-H%). This pattern is a feature of both male and female speakers. For these speakers, the Fo value associated with the final H% boundary in L* H-H% tunes, is usually almost as high as in H* H-H% tunes, resulting in both tune types being perceived as an HRT by transcribers. Effectively for some speakers, there may be some kind of system-internal phonetic difference emerging between the kinds of rise used for questions versus statements, at least with respect to the starting point of the rise. Neither of these rising tunes (L* H-H% versus H* H-H%) would be associated with SSB English declaratives, which suggests that we are still dealing with some kind of semantic difference in tune usage between the two varieties. Speakers of SSB English presumably do not make a linguistic choice to use an HRT with a statement. Some of the Australian English speakers examined here are not only making a choice to use a rising tune versus a falling or fall-rise tune with a statement, they are making a further choice to use a different starting point for the rising tune to differentiate yes/no questions from statements.

Apparently, this potential systematic difference which may be emerging among male and female speakers of Australian English, is not the same as the kinds of differences that are emerging in New Zealand. The question of how the latter may be characterized (i.e. as semantic, systemic, or realizational) is examined in the next section.

14.2.2. *Rises in New Zealand English—a realizational or a systemic difference?*

As mentioned above, like Australian English, the most widely discussed distinguishing feature of New Zealand English (NZE) intonation is the high-rising terminal (e.g. Wells 1982; Britain 1992; Cruttenden 1994). More recent work (e.g. Daly and Warren 2001) has been revisiting the issue of gender differences in intonation patterns in NZE. As well as general features of pitch range and dynamism, this research has been investigating the realization of a range of final rises (not just HRTs) in NZE. A summary of the main findings is outlined below. Initial findings suggest that like the rises in Australian English, it may be necessary to take a closer look at exactly what kind of difference they represent in relation to other varieties of English.

The material referred to here comes from a close replication of the design established for British varieties discussed in further detail in Section 14.3 of this chapter. Initial close scrutiny of data produced in a sentence reading task

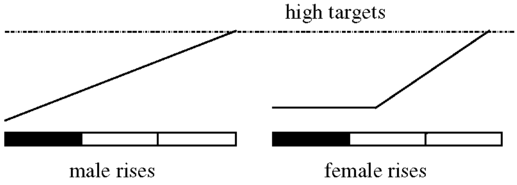


FIGURE 14.4 Schematic representation of male and female rises, showing the interrelationship of rise size, speed, and alignment. The oblongs represent syllables; the filled one is accented. The dotted line indicates the target pitch value for the rise.

revealed a difference in how question rises were being realized. All relevant cases involve a nuclear rise associated with a final di- or trisyllabic word with initial stress. While males tend to start a rise in the accented syllable, the females start it later, in the post-accented syllable. Figure 14.4 is a schematic representation of these two different phonetic rises. It is not clear whether the difference is realizational (Ladd's Type 3 difference), and thus a potential socio-phonetic marker of gender identity; systemic (a Type 2 difference), indicating a difference in the inventories of tunes from which females and males select (so again a potential marker of gender identity); or whether in fact the groups of speakers are simply making different semantic choices from a common pool of tunes.

The differences between the genders are illustrated by Warren and Daly's (2000) data for echo questions. While the late rising pattern is commonly used by both groups, males frequently have an early sharp rise to a high on the accented syllable, giving L+H* H-H%. Females, by contrast, often exhibit a later rise, possibly L* L-H%, but plausibly a realizational variant of L* H-H%. These differences are illustrated in Figures 14.5 and 14.6. In Figure 14.6, the rise has been labelled L+H* H-H% to capture the dynamic nature of the rising tonal movement through the initial primary stressed syllable of 'lilies'. Over the entire set of question sentences, late rises account for 54 per cent of the female questions (with 25 per cent rises on the initial syllable and 21 per cent no rises), but for only 17 per cent of the male questions (59 per cent were initial and 24 per cent had no rises). Informal judgements by native speakers of the dialect suggest that the male L+H* H-H% pattern on echo questions does not mark any particular additional nuances, but that the females L* L-H% or L* H-H% may indicate 'polite insistence/reminder', a meaning also conveyed by this sequence in other varieties of English. This suggests a gender difference in the approach taken to the task, at least as far as these late rises are concerned, rather than either a realizational or a systemic difference. It also suggests, though, that NZE at least as exhibited by the male speakers, has as one of its unmarked forms of question rises a contour similar to the Northern British rise-plateau (see Section 14.3).

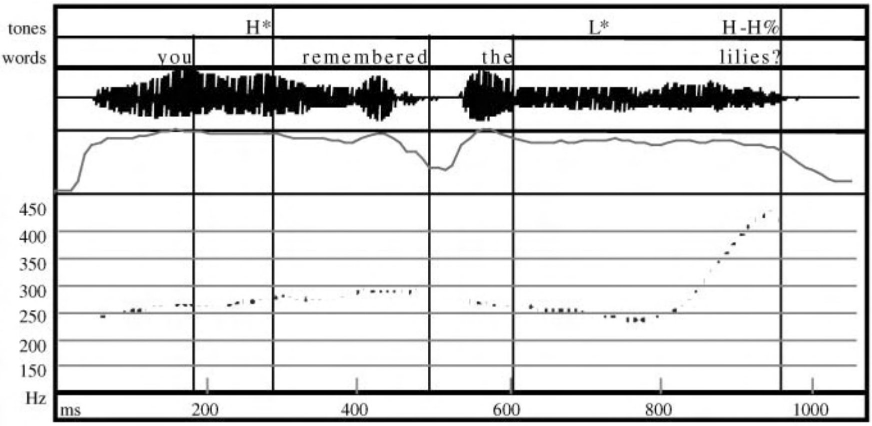


FIGURE 14.5 An actual example of a late-rising tune produced by a female speaker of New Zealand English for the yes/no question 'You remembered the lilies?' The rise has been transcribed here as L* H-H%. Many of these rises can also be transcribed as L* L-H%.

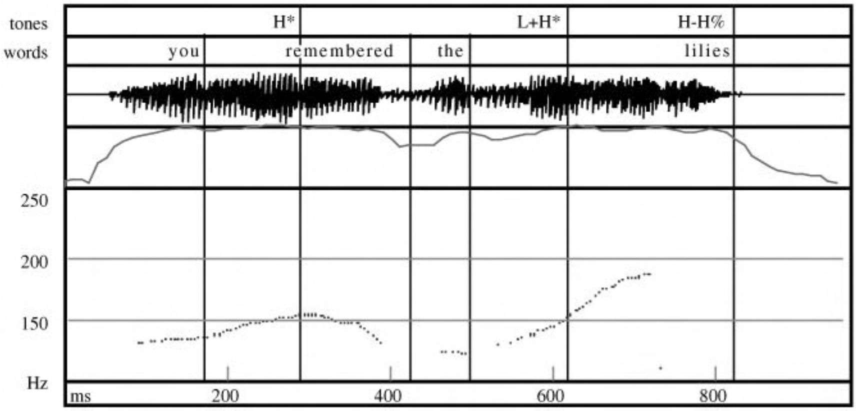


FIGURE 14.6 An example of an early rising tune produced by a male speaker of New Zealand English, transcribed L+H* H-H%.

In the NZE map task data, Warren and Daly distinguished the use of rises in questions from their use on statements, where they fulfil a function frequently associated with HRTs, namely checking that the other participant is following the instructions. Unlike the main trends reported for the Australian MAP data in the preceding section, there is a clear distributional difference between males and females, with the latter having later rises. However, both male and female speakers show a larger proportion of late rises in statements compared with questions. On the basis of rather limited data (in particular, the males ask few questions in this task), it is unclear whether this pattern is consistent enough to constitute a phonological difference between question and statement rises. With other tunes, there is some additional evidence that males and females may be using similar phonological patterns, but aligning them differently. For instance, the sharp nuclear rise before a high boundary (L+H* H-H%) that is frequently used by males in the sentence-reading task turns up in the female data from the map task, but often with a later alignment. We do not believe that this is a contrasting L*+H scooped rise, though further data are needed to verify this.

14.2.3. *Neutralization of intonational contrasts*

According to Wells (1982), an additional source of difference among dialects is the extent to which phonemic contrasts may be neutralized *under certain conditions* (our emphasis). It may be pertinent to extend Ladd's typology to include this possibility. For example, the HRTs of Australian and New Zealand English may represent a form of 'tune' neutralization. In the case of Australian English HRTs, it has been assumed that a particular tune (e.g. H* H-H%) can have two possible semantic interpretations—either signifying a yes/no question or declarative statement. Another way of looking at it is that in the genres where HRTs are most prevalent (e.g. narratives and pragmatically complex texts like map task interactions), the contrast between the high (or low) rising tunes and the 'normal' falling declarative contours may be neutralized in certain contexts. Note however, that one of the 'consensus' interpretations of the HRT by almost all of the earlier studies of this phenomenon in Australian English is that it is often used as a floor-holding device. This suggests that an alternative form of 'functional' neutralization may be taking place. The contrast between the fall-rise tune (e.g. H* L-H%), or 'continuation' contour and high rising tune may be suspended. Critically, speakers who employ HRTs also use falling tunes and fall-rise tunes. However, in certain pragmatic contexts an HRT is favoured. This analysis of tune usage is highly speculative and needs further investigation across the HRT-using varieties. Nevertheless, neutralization of 'phonological' contrasts may be a useful addition to Ladd's typology of intonational differences when comparing tune use among varieties of a language.

In the next sections, we turn our attention to the rising tunes of two Urban Northern British varieties. We also consider the potential problems these varieties present for intonation transcription systems built on the notion of a 'standard' variety.

14.3. INTONATION IN THE BRITISH ISLES

14.3.1. *Rises in Belfast English and Glasgow English*

The 'rise-plateau' and 'rise-plateau-slump' patterns prevailing in many northern varieties of British English have attracted some attention in the literature (Cruttenden 1994; Ladd 1996; Mayo, Aylett, and Ladd 1997; Nolan and Grabe 1997; Grabe 1998*a*). The considerable level of intonational variation in British English is being investigated in a longer-term project at the University of Cambridge—Intonational Variation in English (IViE).[2] Results from the IViE project are available in Grabe *et al.* (1998), Evans and Grabe (1999), Nolan and Farrar (1999), and Grabe, Post, Nolan, and Farrar (2000).

The 'Urban Northern British' (UNB) rises (Cruttenden 1994) are cited as the classic example of a systemic difference or Type 2 difference by Ladd (1996) where the intonational difference lies in the inventory of phonologically distinct tune types, irrespective of semantic differences. The rises represent a typical declarative tune in many of the Northern British varieties and are therefore quite different from the high rises discussed in the preceding sections. In the original ToBI annotation system for American English, the combination L* H-L% transcribes a rise-plateau. The H- phrase tone 'upsteps' the final L% boundary tone. This tone combination is not, therefore, available to transcribe a rise-plateau-slump in the Northern British varieties. Thus, the rise-plateau-slump represents a Type 3 contrast, where the H-L% transcription has quite a different phonetic realization in Glasgow or Belfast English compared to standard American or General Australian English. The transcription solution offered in Glasgow ToBI or GlaToBI (Mayo 1996; Mayo *et al.* 1997) involves the removal of the upstep rule after an H- phrase tone; the rise plateau is transcribed as L*H H-H% and the rise-plateau-slump as L*H H-L%. This solution may not produce transcriptions, which are comparable across different varieties of English (i.e. L*H H-H% transcribes a rise-plateau-rise in Southern British English and a rise-plateau in Glasgow English), but it works for Glasgow English if it is assumed that this variety does not have high rising tunes, like SSB English or Australian and New Zealand English.

[2] The aim of the project (Intonational Variation in English, or IViE, Economic and Social Sciences Research Council award R000237145 to E. Grabe and F. Nolan, 1997–2002) is to make available on CD-ROM an orthographically transcribed and prosodically-labelled database containing speech samples from eight varieties of British English. The varieties of English investigated are spoken in London, Cambridge, Leeds, Liverpool, Newcastle, Belfast, and Dublin. All speakers performed five tasks producing read, semi-spontaneous, and interactive speech data. The speech samples are transcribed using the IViE system for prosodic labelling.

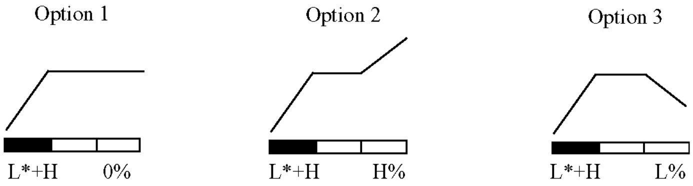


FIGURE 14.7 Three boundary options in Belfast English.

There is a variety of English, however, which unlike Glasgow English clearly exhibits three boundary options: Belfast English. Data from the IViE corpus (see Section 14.3.2(i) show that after an L*H sequence of targets, Belfast speakers produced predominantly high plateaus in read speech but they also produced rises and falls, i.e. with pitch continuing to rise, or then fall after the L*H sequence (Grabe *et al.* 2000). Thus, when transcribing Belfast English, suspending the upstep rule is not always a satisfactory solution.

The solution offered in the IViE system is the following. At least in principle, it is assumed that speakers have three options at every phrase boundary. The implementation of the options is variety-specific; in Cambridge English, for instance, speakers have two options, in Belfast English, they have three. In the absence of a stressed syllable, speakers may raise pitch, lower pitch, or leave matters as they are. Rising pitch is transcribed as H%, falling pitch as L% and no change is transcribed as 0% (Grabe 1998*a*). Figure 14.7 illustrates the boundary options in Belfast English and gives the corresponding transcriptions.

14.3.2. *Some representational issues in transcribing different varieties*

Due to the systemic and realizational differences between rising tunes in Glasgow and Belfast English versus American, Australian, or New Zealand English discussed in the previous sections, the issue of representation has needed to be addressed by researchers working on these varieties. Tagging conventions based on autosegmental metrical treatments of English intonation (e.g. Pierrehumbert 1980) have either been modified (e.g. the removal of upstep) or different autosegmental-metrical models (e.g. Gussenhoven 1984; Grabe 1998*a*) have provided the basis for intonational tags to describe rising boundary configurations in UNB and other varieties, as well as other intonational phenomena.

So far it has been assumed that the main source of intonational variation between Australian English and American or British English is to do with tune choice and/ or differences in the phonetic realization of phonological categories, which are shared among the varieties. In other words, the variation is not necessarily due to differences in phonological inventories. The application of the American English ToBI annotation conventions to Australian English is relatively uncontroversial largely because the intonational phonology of these varieties is very similar (Fletcher and Harrington 2001). The conventions sanctioned by the tone and break index tiers provide sufficient coverage of the major intonation patterns and higher-level prosodic characteristics of Australian English. The direct application of a version of the American ToBI tagging system (or any other annotation system for SSB, American, or Australian English) to New Zealand English may not be as straightforward. The relevant phonological categories of New Zealand English intonation need to be established before adopting a labelling strategy that is identical to other varieties. At this stage no assumption has been made that the same annotation system (ToBI or otherwise) will automatically hold across both Australian English and New Zealand English. More rigorous quantitative analysis is needed before the relevant phonological contrasts are established for NZE. This needs to be done before a particular set of annotation conventions can be adopted to represent these phonological contrasts. On the other hand, due to the well-attested existence of intonational differences among varieties of British English, it is of no surprise that at least two annotation systems have been developed to reflect these differences. Some of the labelling conventions adopted by the IViE system and the GlaToBI annotation system have already been outlined in Section 14.3.1. Further details of these systems will now be discussed.

(i) *IViE*: the IViE system is modelled on the original ToBI conventions for American English, but incorporates two major changes (Grabe *et al.* 1998). The first involves the tonal inventory, and the second involves the number of tiers available to the transcriber. Changes to the tonal inventory were made to allow for comparable transcriptions of more than one variety of English in a single transcription system; unlike the original ToBI conventions for American English which accounts for one particular variety of English (i.e. the so-called 'standard' American English variety), IViE offers a pool of labelling options from which transcribers can choose a subset of labels for each variety they investigate. The IViE labels themselves are based on phonological analyses of English intonation by Gussenhoven (1984) and Grabe (1998*a*).

Secondly, the IViE system offers two new tiers, the rhythmic and the pitch movement tier. The addition of the two extra tiers results in the following 5-tier system:

(1) orthographic tier
(2) rhythmic tier

(3) pitch movement tier
(4) phonological tier
(5) miscellaneous tier.

Note that IViE does not have a break index tier, because the system does not deal with different degrees of disassociation between words within intonation phrases or different degrees of boundary strength. The rhythmic and pitch movement tiers are intended to increase the transparency and replicability of the labels on the (phonological) tone tier. In essence, they permit a step-by-step breakdown of the process, which leads to a specific tonal transcription. In English, this process begins with the identification of rhythmically prominent (stressed) syllables because the pitch movements transcribed on the tone tier are anchored to these syllables. In IViE, this identification process is overt, rather than implicit; a rhythmic tier has been added on which the location of rhythmically prominent (i.e. potentially accentable) syllables is transcribed, by aligning the label 'P' for 'prominence' with the relevant vowel. The second step in the prosodic labelling procedure involves the identification of rhythmically prominent syllables which are not only stressed but accented, that is, associated with pitch movement (note that some prominent syllables may not be accompanied by pitch movement). Accentedness is established via inspection of the fundamental frequency trace and careful listening. The pitch movement surrounding the stressed syllable (if any) is then transcribed on the 'pitch movement tier'. Note that the pitch movement tier has heuristic rather than linguistic status; it allows labellers to make a record of the impression of a particular pitch movement which, combined with other information, leads them to assign phonological labels to a contour at a later stage. The pitch movement tier makes that decision-making process accessible to users of IViE transcriptions.

Secondly, the pitch movement tier provides information about accent realization. In varieties of British English, the realization of a pitch accent varies with (a) the segmental structure it is associated with and (b) the location of pitch accent within an IP (Grabe 1998*b*; Nolan and Farrar 1999; Grabe *et al.* 2000). On the pitch movement tier, the labeller provides information about the realization of pitch accents. The surface realization of a particular accent is transcribed within pitch movement *Implementation Domains* or IDs. Relevant landmarks within the ID are (1) the preaccentual syllable, (2) the accented syllable and (3) any following unaccented syllables (if any) up to the next accented syllable. Put simply, an ID consists of the preaccentual syllable and the following 'accent foot'. The examples in Figures 14.8 and 14.9 are intended to give a flavour of the labels on the pitch movement tier. Labels available are

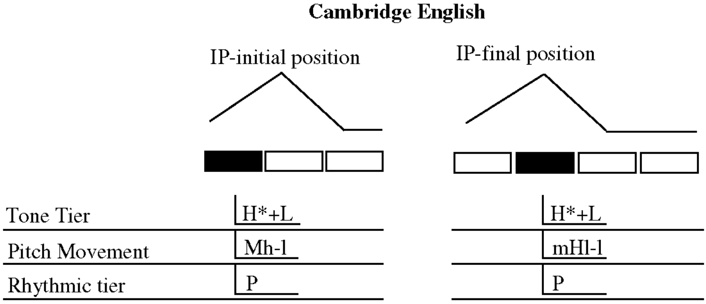


FIGURE 14.8 Realization of H*+L in IP-initial and IP-final position.

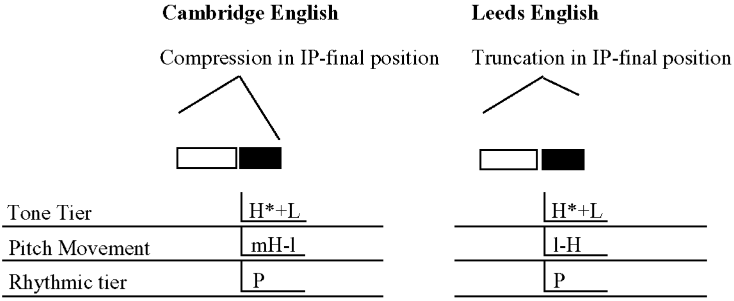


FIGURE 14.9 Cross-varietal differences in the realization of H*+L.

h(igh), l(ow) and m(mid), and they are transcribed relatively to each other. Capital letters indicate a pitch level accompanying a stressed syllable.

Figure 14.8 shows that in Cambridge English, H*+L in IP-final position is realized with high pitch on the stressed syllable (transcribed as H), low pitch on the following syllables (l), and mid-pitch (m) on the preaccentual syllable (Grabe 1998*a*). In IP-initial H*+L, the pitch peak is frequently delayed beyond the accented syllable, and we find mid-pitch on the stressed syllable (M), followed by a pitch peak on the following unaccented syllable (h). The syllable or syllables immediately after the peak are low (Nolan and Farrar 1999). A comparable effect can be observed in Newcastle, and in Leeds English.

Figure 14.9 illustrates cross-varietal realizational differences conditioned by segmental structure. The figure shows that in Cambridge English, on very short syllables with little voicing, H*+L is realized as a steep fall in Fo (compression). In Leeds English, H*+L on the same syllable accent is realized differently; instead of a steep fall, we see a very shallow fall or a level, and we hear high pitch rather than falling pitch (cf. Grabe 1998*b*, Grabe *et al.* 2000). The pitch movement tier allows us to capture this difference. The assumption is that it is possible to establish a one-to-many mapping between a specific phonological label and a finite set of pitch movement labels. The relationship between the pitch movement tier and the ensuing phonological labels makes this mapping explicit and provides one method for determining whether intonational differences among varieties are realizational rather than systemic, according to Ladd's typology.

(ii) *GlaToBI:* the GlaToBI annotation system was designed to annotate intonation and prosody for one specific variety, namely Glasgow English, unlike the IViE system that aims to capture similar and dissimilar intonational phenomena across a number of varieties. GlaToBI also includes a Break Index tier, like other 'ToBIs'. Modifications have been made to the original tone tier of ToBI to represent the characteristic tunes of 'Standard' Western Scottish English, in particular the variety spoken in Glasgow. Similar to many recently developed annotation systems, the development of GlaToBI was intrinsically linked with the aim of performing large scale intonational analysis of a substantial digital speech corpus, in this case, the HCRC Map corpus (Anderson *et al.* 1991).

The tone and break index tiers are the central components of GlaToBI. The Break Index tier is essentially unchanged from American English ToBI. The indices range from 0 to 4 with the latter representing the highest-level intonational constituent, the intonational phrase (see Mayo 1996, for further details on other tiers). There are at least two crucial differences between the ToBI tone labelling conventions used for Glasgow English, and those adopted for American or Australian English. The first is the elimination of the contrast between rising and scooped accents, L+H* and L*+ H, and second, already discussed in Section 14.3.1, is the removal of the upstep rule after an H- phrase tone to take account of the rise-plateau-slump. We mentioned that one of the implications of this is that the L*(+)H H-L% transcription will account for a very different tune in American or Australian English than in Glasgow English.

The elimination of the contrast between the two-bitonal accents (L+H* and L*+ H) is another potential systemic or Type 2 difference. The accent type labelled by Mayo, Aylett and Ladd as L*H indicates that the accent is associated with a rising tune. They also claim that the * does not denote that either the H or L tone is phonetically anchored to the stressed syllable, but rather the *movement* from one to the other is what is observed through the associated stressed syllable. In other words, the temporal alignment of the L or H tone with the stressed syllable, critical to the contrast between the L*+ Hand L+H* of other English varieties, is not a feature of intonation patterns observed in the Glasgow English MAP data. It is therefore not absolutely necessary to include two different bitonal labels in the GlaToBI tone inventory. Mayo (1996) suggests however, that there may be another kind of rising accent in Glasgow English, which aligns the H tone of the LH configurations with the stressed syllable. She therefore recommends retaining the two bitonal choices L*+H and L+H* in the inventory together with the L*H tag in case alignment proves to be a crucial factor in distinguishing more than one kind of rising accent in Glasgow English.

14.4. CONCLUSION

In this chapter we have presented a brief account of the characteristic high rises of four varieties of English. We have examined how they can be described using a typology of intonational differences proposed by Ladd. Differences that have been classified as 'semantic', such as the use of the high rising tune with statements by speakers of Australian and New Zealand English, may also constitute realizational differences. Furthermore, the phonetic realization of the HRTs and rises in general may not be entirely identical in the two Australasian varieties. There also appears to be gender-related sources of variation in the phonetic realization of rises in each case. The HRTs of Australasian English are therefore more complex intonationally than previous analyses would suggest, and are not necessarily best characterized as a simple substitution or redeployment of a phonetically identical 'question' intonation to a statement. Careful phonetic analysis of a wider variety of data is clearly necessary to support the preliminary findings reported here. We have also suggested that Ladd's typology might be usefully extended to include 'neutralization' of intonational differences. This is particularly relevant in cases where the intonational inventories of two varieties appear to be identical (i.e. between SSB English and Australian English) but where tune choice is clearly different under certain circumstances or in particular genres (e.g. the use of HRTs in narratives).

Some representational issues that arise from considering differences in tune usage or tune inventory among varieties have also been considered with respect to rising tunes and pitch accent realization. On the one hand, it is clear that certain varieties present a more or less straightforward case when it comes to issues of annotation (e.g. the application of American ToB! conventions to Australian English) compared to others (e.g. Glasgow English, Belfast English). This is not necessarily because the original American English ToB! conventions were designed to be 'pan-dialectal' in nature, rather it is more a question of the similar intonational inventories between Australian and American English. There has also been a reasonable amount of discussion in recent years as to whether it is desirable to have the same transcribed tone sequence represent two radically different tunes (e.g. Ladd 1996; Nolan and Grabe 1997; Grabe *1998a;* Grabe *et al.* 1998) across dialects or varieties. In addressing this concern, it is clear that one needs to take into account the major aims of a particular set of annotation conventions when considering representational issues. In the case of IViE, for example, one of the main goals of the system is to provide sufficient coverage of intonational phenomena across a very large corpus comprising several, very different, varieties of British English. An alternative approach is adopted by the developers of GlaToBI whereby a set of specific annotation conventions have been devised to capture the salient intonational events of one specific variety. Both approaches are united, however, in showing that the phonological and phonetic conventions of a so-called 'standard' are not always applicable to all varieties of a language. The example of New Zealand English intonation is also very important here. We noted earlier that the phonological contrasts of this variety (and its various sociolects) need to be established before adopting a rigid set of annotation conventions for a so-called standard variety of New Zealand English.

Documenting intonational variation among speech varieties is of great interest to many researchers of languages other than English. For example, the large-scale quantitative cross-dialectal studies of Swedish and Dutch dialects (e.g. Bruce *et al.* 1999; Bruce this volume Ch. IS; Gussenhoven and van der Vliet 1999) stress the importance of examining intonational variation among dialects of a language. For example, Bruce *et al.* state that this is necessary in order to contribute to the 'definition of criteria on which phonetic and phonological typologies can be based' (Bruce *et al.:* 321). Further large-scale analysis of gender-based or other sociolectal differences within a variety also needs to be undertaken before we can construct new or augment existing phonetic and phonological models and typologies. The research currently being undertaken by Daly and Warren on New Zealand English and the analysis of pitch realization by Latina girls in California by Jannedy and Mendoza-Denton (1999) are further examples of the kind of sociolectal study that can contribute to our understanding of the relationship between an intonational model based on a 'standard', and the type and range of variation that either can, or cannot be accommodated within that model.

REFERENCES

ANDERSON, A., BADER, M., BARD, E., BOYLE, E., DOHERTY, G., GARROD,S., ISARD, 5., KOWT KO, J., McALLISTER, J., SOTILLO, C., THOMPSON, H., and WEINERT, R. (1991), 'The HCRC Map Task Corpus', *Language and Speech,* 34/4: 351-66.

BRITAIN, D. (1992), 'Linguistic Change in Intonation: The Use of High Rising Terminals in New Zealand English', *Language Variation and Change,* 4: 77-104.

BRUCE, G. (this volume Ch. 15) 'Intonational prominence in varieties of Swedish revisited'.

— —, ELERT, C., ENGSTRAND, 0., and WRETLING, P. (1999), 'Phonetics and Phonology of the Swedish Dialects-A Project Presentation and a Database Presenter', in *Proceedings of the Fourteenth International Congress of Phonetic Sciences* (San Francisco), 321-4.

CRUTTENDEN, A. (1994), 'Rises in English', in J. Windsor-Lewis (ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J.* D. *O'Connor* (London: Routledge), 155–73.

DALY, N., and WARREN, P. (2001), 'Pitching it Differently in New Zealand English: Speaker Sex and Intonation Patterns', *Journal of Sociolinguistics,* 5/1: 85-96.

EVANS, B., and GRABE, E. (1999), 'Connected Speech Processes in Intonation', in *Proceedings of the Fourteenth International Congress of Phonetic Sciences* (San Francisco), Vol.1: 33-6.

FLETCHER, J., and HARRINGTON, J. (2001), 'High-Rising Terminals and Fall-rise Tunes in Australian English', *Phonetica,* 58: 215-29.

GRABE, E. *(1998a), Comparative Intonational Phonology: English and German,* MPI Series 7 (Nijmegen, The Netherlands).

— — *(1998b),* 'Pitch Accent Realisation in English and German', *Journal of Phonetics,* 26: 129-44.

— —, NOLAN, F., and FARRAR, K. (1998), 'IViE-A Comparative Transcription System for Intonational Variation in English', in *Proceedings of the 5th Conference on Spoken Language Processing (ICSLP)* (Sydney, Australia), 1259-62.

— —, POST, B., NOLAN, F., and FARRAR, K. J. (2000), 'Pitch Accent Realisation in Four Varieties of British English', *Journal of Phonetics,* 28: 161-85.

GUSSENHOVEN, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

— —, and VAN DER VLIET, P. (1999), 'The Phonology of Tone and Intonation in the Dutch Dialect of Venlo', *Journal of Linguistics,* 35: 199-235.

GUY, G., and VONWILLER, J. (1984), 'The Meaning of an Intonation in Australian English', *Australian Journal of Linguistics,* 4: 1–17.

— —, --(1989), 'The High Rise Tones in Australian English', in P. Collins and D. Blair (eds.), *Australian English* (St Lucia: UQP), 21-33.

HALLIDAY, M. A. K. (1967), *Intonation and Grammar in British English* (The Hague: Mouton).

HORVATH, B. (1985), *Variation in Australian English: The Sociolects of Sydney* (Cambridge: Cambridge University Press).

JANNEDY, S., and MENDOZA-DENTON, N. (1999), 'Low Pitch in the Linguistic Performance of California Latina Gang Girls', *Perceiving and Performing Gender Seminar,* University of Kiel, Germany, 12-14 November.

LADD, D. R. (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

MAYO, C. J. (1996), 'Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI', thesis, University of Edinburgh.

—— , AYLETT, M., and LADD D. R. (1997), 'Prosodic Transcription of Glasgow English: An Evaluation Study of GlaToBI', in A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.), *Proceedings of the ESCA Tutorial and Workshop on Intonation: Theory, Models, and Applications* (Athens, Greece), 231-4.

MILLAR, J., VONWILLER, J., HARRINGTON, J., and DERMODY, P. (1994), 'The Australian National Database of Spoken Language', in *Proceedings of ICASSP-94* (Banff, Canada), 197-100.

MITCHELL, A., and DELBRIDGE, A. (1965), *The Pronunciation of English in Australia* (Sydney: Angus and Robertson).

NOLAN, F., and FARRAR, K. (1999), 'Timing oHo Peaks and Peak Lag', in *Proceedings of the International Congress of Phonetic Sciences* (San Francisco), 961-4.

—— , and GRABE, E. (1997), 'Can ToB! Transcribe Intonational Variation in the British Isles?', in A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.), *Proceedings of the ESCA Tutorial and Research Workshop on Intonation: Theory, Models and Applications* (Athens, Greece), 259-62.

PIERREHUMBERT, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

PITRELLI, J., BECKMAN, M., and HIRSCHBERG, J. (1994), 'Evaluation of Prosodic Transcription Labeling Reliability in the ToB! Framework', in *Proceedings of 1994 International Conference on Spoken Language Processing* (Yokohama, Japan), 1: 123-6.

WARREN, P., and BRITAIN, D. (2000), 'Intonation and Prosody in New Zealand English', in A. Bell and K. Kuiper (eds.), *New Zealand English* (Wellington: Victoria University Press), 146-72.

—— , and DALY, N. (2000), 'Sex as a Factor in Rises in New Zealand English', in J. Holmes (ed.), *Gendered Speech in Social Context: Perspectives from Gown to Town* (Wellington: Victoria University Press), 99-115.

WELLS, J. (1982), *Accents of English* 1: *An Introduction* (Cambridge: Cambridge University Press).

15

Intonational Prominence in Varieties of Swedish Revisited

Gösta Bruce

15.1. INTRODUCTION

In reports on Swedish intonation the description is not seldom confined to one variety of the language, Standard Swedish (Stockholm Swedish) (cf. e.g. Bruce 1977, 1987; Engstrand 1995, 1997). It should be emphasized, however, that there is considerable variation among Swedish dialects in terms of intonation and accentuation (cf. e.g. Meyer 1937, 1954; Öhman 1967; Gårding and Lindblad 1973; Gårding 1977; Bruce and Gårding 1978). This intonational variation was the object of study in the research project Swedish Prosody conducted in Lund in the 1970s and supported by the Swedish Humanistic Research Council (cf. Gårding 1982). The present contribution recapitulates and puts emphasis on exactly this variation among varieties of Swedish.

There are at least two reasons for me to take up this topic again. One is the new SweDia 2000 research project, the Phonetics and Phonology of the Swedish dialects around the year 2000. This project, funded by the Bank of Sweden Cultural Foundation, is a cooperation between departments of phonetics in Lund, Stockholm, and Umeå with the aim of data collection and phonetic analysis of more than 100 varieties of Swedish (Engstrand *et al.* 1997). One specific area of interest within this general phonetics project will be prosody and prosodic dialect typology (Bruce *et al.* 1999; Bruce forthcoming). Another reason is the recent reanalysis of the Scandinavian pitch accent system proposed by Tomas Riad (Riad 1998). This typology focuses particularly on the phonological aspects suggesting a set of constraints to describe Scandinavian pitch accents and a ranking of these constraints within the framework of optimality theory.

A basic assumption of mine is that there is in Swedish (as well as in other Germanic languages) a fundamental distinction between rhythmical

	category	unit	domain	correlate
RHYTHMIC	stress	syll	foot	duration
prominence				intensity
				spectrum
INTONATIONAL	accent	foot	word	pitch
prominence	focus	word	phrase	pitch

FIGURE 15.1 Rhythmical and intonational prominence. Model of phonological prominence levels in Swedish.

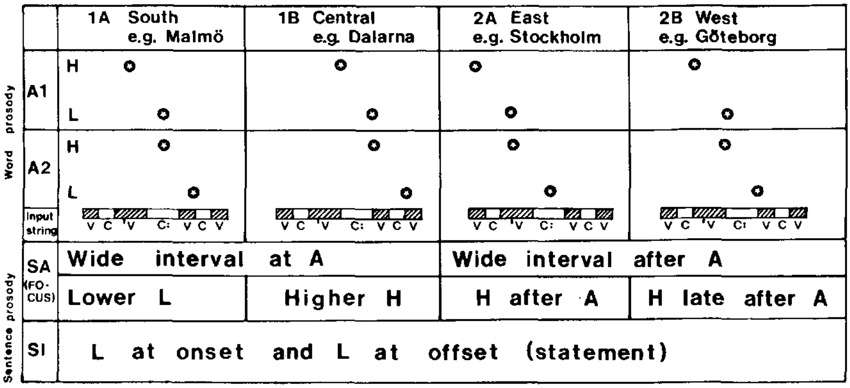
prominence and intonational prominence (see Figure 15.1). Stress is the fundamental category of rhythm. There is an alternation between stressed and unstressed syllables, which forms the basis of the rhythmical structure and the system of prominence levels in Swedish. Stress has a fairly complex phonetic cueing involving relations of duration, intensity, and spectrum. On top of this basic structure of rhythmical prominence, intonational prominence acts as a kind of amplification of certain parts of it and thus as a higher level of prominence. Intonational prominence is cued mainly by pitch. Two hierarchical levels of intonational prominence appear to be relevant in Swedish: accent (non-focal accent) and focus (focal accent). Under each higher level of prominence (accent and focus), a word form has got either of the two word accents, accent I (acute) and accent II (grave). There is no difference in prominence level between the two word accents. This modelling of prominence and stress degrees gives us three phonologically distinct levels of prominence besides the unstressed category. See further Bruce (1977) for a more complete reasoning about the issue of number of prominence levels. In the present paper I will focus particularly on intonational prominence (pitch accents) in Swedish, i.e. on the two levels of intonational prominence called accent (non-focal accent) and focus (focal accent).

Within the project Swedish Prosody, a prosodic dialect typology for Swedish dialects was developed (Bruce and Gårding 1978). The database used in the project contained phonetically balanced utterances with a systematic variation of (besides accent I/accent II) placement of focus, final/non-final phrase position and simplex/compound words, i.e. typical laboratory speech data. A main feature of this typology was the timing of the pitch accent gesture as critical for the distinction between accent I and accent II. Another critical feature was the pitch realization of focus. Yet another feature, and a third characteristic of the typology was the recognition of the pitch patterns of compounds as a criterion of prosodic dialect type.

The main body of my paper will be divided accordingly into three sections: timing of pitch accent gestures, pitch realization of focus, and pitch patterns of compounds. The general aim of my contribution is a critical assessment of the analysis of intonational prominence in the Scandinavian dialects and in particular varieties of Swedish.

15.2. TIMING OF PITCH ACCENT GESTURES

In our modelling of intonation (Bruce and Gårding 1978), the distinction between accent I and accent II (in the varieties of Swedish that have the difference) was a difference in the timing of the pitch accent gesture in relation to the segmentals and in particular to the stressed syllable. The pitch accent gesture was modelled in terms of H(igh) and L(ow) turning points. The H+L gesture for accent I appeared to be earlier than for accent II independent of dialect. The timing of the H+L gesture was a relevant parameter not only for the word accent distinction (accent I/accent II) but also for each of the word accents in an inter-dialectal comparison. In this way there appeared to be a distinct order of the four dialect types recognized in the typology from early to late timing: EAST (Stockholm), WEST (Gothenburg), SOUTH (Malmö), CENTRAL (Dalarna) (see Figures 15.2 and 15.3). There are, however, a few problems connected with our modelling of pitch accent timing. One general problem is that in our prosodic dialect



		1A South e.g. Malmö	1B Central e.g. Dalarna	2A East e.g. Stockholm	2B West e.g. Göteborg
Word prosody	A1	H ○ L ○	○ ○	○ ○	○ ○
	A2	H ○ L ○	○ ○	○ ○	○ ○
	Input string	V C ˈV C: V C V	V C ˈV C: V C V	V C ˈV C: V C V	V C ˈV C: V C V
Sentence prosody	SA	Wide interval at A		Wide interval after A	
	(FOCUS)	Lower L	Higher H	H after A	H late after A
	SI	L at onset and L at offset (statement)			

FIGURE 15.2 Swedish intonation model: linguistic components. Dialectal representations of word prosody (accent I [A1] and accent II [A2]), sentence prosody (sentence accent [SA, focus] and sentence intonation [SI]) for four prosodic dialect types (from Bruce and Gårding 1978).

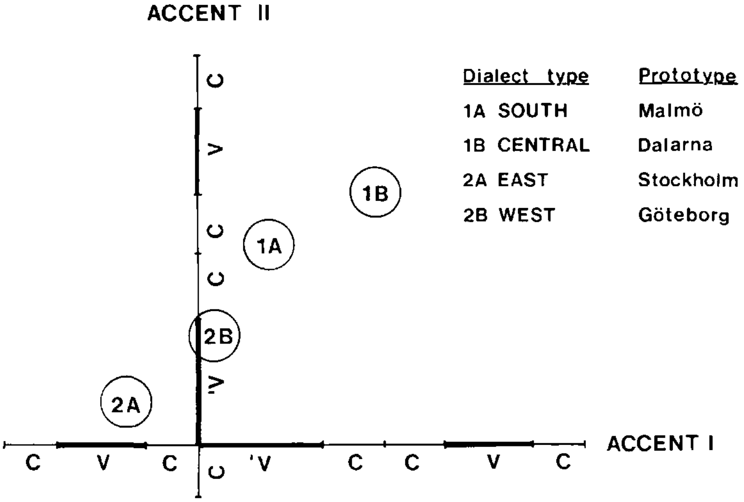


FIGURE 15.3 Swedish word accents in an inter-dialectal comparison. Schematic diagram showing the timing of the H(igh) turning point in relation to the segmentals for Swedish word accents, accent I (x-axis) and accent II (y-axis), for four prosodic dialect types.

typology we confined the modelling of the basic pitch accent gesture to a combination of H and L turning points. A consequence of this particular modelling was that in cases where the H+L gesture has a very late timing, e.g. Dalarna accent II, none of the turning points is located in the stressed syllable. As the association of a pitch accent gesture is assumed to be with the stressed syllable, at least one of the turning points (H or L) would be expected to be tied to the stress. Another general problem to be dealt with is that no explicit restrictions on the number of available alignments of the H+L gesture in relation to the segmentals were assumed. As noted above, the specific temporal alignment of the H+L sequence with the segmentals varied with accent type (accent I or accent II) and dialect (four prosodic dialect types) but was assumed to be free with no restrictions on the number of possible timings. This presents a problem both for phonology, which is typically assumed to be categorical, and for phonetics, as we may expect restrictions on human processing of pitch accent timing. In the larynx model proposed by Öhman (1967) a continuously variable timing of the word accents was also implied. This was demonstrated in the arrangement of the Scandinavian dialects by Öhman in a so-called Scandinavian accent orbit. A more specific problem with our modelling of pitch accents is related to the representation of accent I in EAST (Stockholm) Swedish as H+L*, particularly the status of the leading H.

15.2.1. *Phonological representation of pitch accent timing*

A first general point is that instead of having only a H+L gesture, as in our modelling of pitch accent timing in varieties of Swedish, we would assume a basic L+H+L gesture which can have a number of different timings in relation to rhythmical line-up points. Depending on dialect and word accent, different parts of the L+H+L gesture will hit the stressed syllable. This would solve one of the problems referred to in the preceding section, namely the late timing of Dalarna accent II. But there are further problems to be tackled. In my own reanalysis of the Bruce and Gårding (1978) model I used starring of tones in accordance with current usage in autosegmental phonology to represent the distinction between the two word accents in EAST (Stockholm) Swedish (Bruce 1987). This appears to be straightforward for Stockholm Swedish, where either the L (for accent I) or the H (for accent II) of the H+L gesture is specifically associated with the stress, and with the other tone as merely leading or trailing respectively. But for the description of pitch accent timing in other varieties of Swedish this specific association of one of the tones of the bitonal gesture is not as evident at all. The number of available associations of a (bitonal) pitch accent gesture is highly limited by accepting starring of tones in this way. Even if we assume that both H+L and L+H can represent a pitch accent, only four different associations of a bitonal gesture can be distinguished from early to late timing: H+L*, H*+L, L+H*, L*+H.

Recently the general validity of an obligatory starred tone of a pitch accent has been questioned. Examples are taken from accentuation in Modern Greek (Arvaniti *et al.* 1998) and Glaswegian English (Ladd 1996), where it seems clear that it can be the whole gesture that is synchronized with the stressed syllable rather than only one of the tones of a bitonal pitch accent. It also seems reasonable to assume that some of the accents of the Swedish prosodic dialect types are better described by taking the whole tonal gesture—either the H+L or the L+H—to be associated with the stress, and not only one of the tones. This will be discussed below.

One interpretation of this is that the current usage of the phonological symbolization of pitch accent timing is not sufficient to express observable timing distinctions. Another way of handling this problem is to conceive of a distinction between association (digital) and temporal alignment (analogue) and thus between phonology and phonetics respectively, as has been suggested by Ladd (1983). But before accepting that the particular timing of a pitch accent is merely (or primarily) a phonetic issue, I am inclined to explore the possibilities of expressing distinctions in pitch accent timing within the phonological framework. Let us assume a fairly concrete phonology where Hs and Ls are conceived of as tonal turning points, and distinctions in pitch accent timing are expressed as directly as possible.

The following reanalysis will be presented as a preliminary reinterpretation of the original Bruce and Gårding (1978) analysis of Swedish accentuation. My starting point here is a framework with a basic L+H+L accentual gesture, where typically a part of it—the L+H or H+L part—can be synchronized with a rhythmical line up point in a number of ways. A further assumption is that either of the tones (H or L) of a bitonal accent is synchronized with the stress or both of them are, i.e. the whole rise (L+H) or fall (H+L). The presence of a star on one of the tones then signals the specific alignment of that tone with the stress, while starring both tones means that the whole gesture (H+L or L+H) is aligned with the stress. This convention is proposed just for simplicity reasons. The issue here is what can be expressed within the given framework, not how this should be optimally symbolized. Thus for a basic L+H+L accentual gesture we can identify the following available timings for bitonal accents from early to late: H+L*, H*+L*, H*+L, L+H*, L*+H*, L*+H (see Figure 15.4). For the four Swedish prosodic dialect types, this will give us the following, highly symmetrical representations of non-focal accent I and accent II.

	East	West	South	Central
accent I	H+L*	H*+L*	H*+L	L+H*
accent II	H*+L	L+H*	L*+H*	L*+H

It should be emphasized that these are the basic non-focal accents. More complex pitch patterns will result, when focal accents and boundary tones are added.

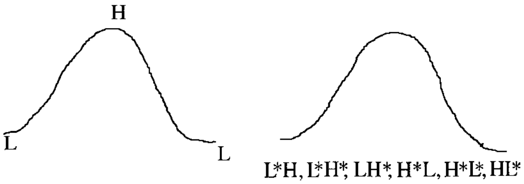


FIGURE 15.4 Temporal alignment of tones. The basic L+H+L accentual gesture (left) and available timings of the L+H or H+L part of it (right).

15.2.2. *Phonetic restrictions on pitch accent timing*

Generally, there are likely to be restrictions on human capacity of processing pitch accent timing. This seems to be true of both production and perception. House (1990) has argued for one such processing constraint from a perceptual viewpoint. His idea is that pitch sensitivity is not constant over time but varies with signal complexity. The human auditory system has to deal with both spectral and tonal information at the same time. For example at the transition from a consonant to a vowel (as in a stressed syllable), a phase of maximal spectral change occurs. This will decrease pitch sensitivity at this point in time and the capacity of the perceptual system to resolve tonal movement. In the medial part of a vowel we have more of spectral stability which increases pitch sensitivity, as the perceptual mechanism is less constrained to resolve tonal movement. At the transition of the vowel and the next consonant there is another phase of spectral change, which will cause pitch sensitivity to decrease temporarily. This means that there is likely to be an uneven distribution of available pitch accent timings in relation to different parts of a stressed CVC syllable.

Perceptual testing of German pitch accent types by Kohler (1987) has shown there to be three distinct categories of Fo peak timing referred to as early, medial, and late and having distinct communicative functions. These three different pitch accent timings imply established fact, new fact and contrastive fact respectively. The early timing of a pitch accent has an Fo peak location occurring even before the CV-boundary of the stressed syllable and a fall through the vowel. The medial timing has a peak in the centre of the stressed vowel, and the late timing has a peak at the end of the stressed vowel or even beyond the stressed syllable preceded by a rise. These three pitch accent categories have been referred to in the terminology used by Pierrehumbert and Steele (1989) as H+L*, L+H*, and L*+H respectively.

The pitch accent types in American English differing in tonal alignment have been described as H*, L+H*, and L*+H by Pierrehumbert and Steele (1989). The H* pitch accent simply adds information, while the two other pitch accents imply a comparison to alternatives, the L+H* conveying certainty and the L*+H uncertainty. The two latter pitch accent types have been tested in a perceptual experiment and shown to be categorically distinct. As the methodology used was the imitation by the subjects of a continuum of synthesized stimuli differing systematically in tonal alignment, production was also involved.

Even if these two languages do not have lexically contrastive pitch accents, it is generally clear from the perceptual testing of pitch accent types in

German and English that we have to do with distinct tonal categories and not with a continuum of possibilities. While the two pitch accent types representing alignment later in a stressed syllable (L+H* and L*+H) appear to be the same in German and English, it is evident, however, that the early timing pitch accent of German (H+L*) is distinct in terms of alignment from the early timing pitch accent of American English (H*).

The Swedish word accents have also been tested perceptually. For East (Stockholm) Swedish a critical difference in the timing of the same pitch accent contour has shown there to be a categorical distinction between accent I and accent II, expressed as H+L* vs. H*+L respectively (Bruce 1977, 1987).

Also the declarative vs. interrogative intonation in Neapolitan Italian has been tested in perceptual experiments (D'Imperio and House 1997). The perceptual testing clearly shows that by shifting an Fo peak through a stressed vowel and thereby creating a basically falling or rising contour a categorical distinction between statements and questions is achieved. The labelling of these pitch accent categories suggested by the authors is HL—either H+L* (broad focus) or H*+L (narrow focus)—for the declarative pitch accent, and L+H* for the interrogative pitch accent.

The above experimentation on pitch accent timing in different languages clearly demonstrates that we should assume a limited number of pitch accent timings to be exploited in human language. But exactly how we should interpret the relationship between phonological representation and phonetic actualization is still an open question.

15.2.3. *Pitch characteristics of accent I in EAST (Stockholm) Swedish*

One more specific point to discuss is the analysis of the pitch accent gesture of accent I in EAST (Stockholm) Swedish, the representation of the non-focal variant as H+L* and of the focal variant as H+L* H^{-}. The interpretation of the word accent distinction in Stockholm Swedish as one of timing of the same basic pitch accent gesture H+L was proposed in my thesis (Bruce 1977). This analysis, in particular the positing of a leading H in a H+L gesture for accent I was challenged from the very beginning. The traditional interpretation of the accent I/accent II distinction in Stockholm Swedish before my new analysis, was to describe the citation forms of the word accents—typically disyllabic words stressed on the first syllable—as a difference in the number of tone peaks. What we would now call focal and phrase-final (terminal juncture) accent I was described as a rise-fall (one pitch peak) and the corresponding accent II as a fall-rise-fall (two pitch peaks). While my decomposition of the pitch accent contours of the word accents into a focal accent part and a terminal juncture part seems to have been generally accepted, my positing of a separate accent I gesture (before the focal accent gesture) parallel to the one recognized for accent II was met with scepticism and was not as readily accepted by Scandinavian colleagues. The pieces of evidence presented against positing a specific early peak for accent I proper, i.e. in my analysis the H of the H+L* gesture, were the following.

Firstly, it has been argued that this H appears to be absent in some contexts, most evidently in a phrase or utterance initial position, and even in some other contexts when the accent I word occurs in a focused position. Secondly, it has been pointed out that the accent I pitch pattern is generally more variable than the corresponding accent II pitch pattern, which appears to be highly stable across context variation (Engstrand 1995, 1997). Thirdly, and in support of the second point, reference has been made to the phonological status of the word accent opposition in Swedish. Traditionally it has been analysed as a privative opposition, where accent II is the marked member (cf. Elert 1964). This is also the position in Riad (1998). In addition to phonological and morphological evidence in favour of this interpretation, the phonetic evidence above appears to tie in with the word accent opposition being privative and not equipollent, as more or less implied by the Bruce (1977) and Bruce and Gårding (1978) analysis. Alternatives to the bitonal and timing analysis in our modelling of accent I/accent II have been to represent the distinction as either a monotonal opposition L versus H (Withgott and Halvorsen 1988 for Norwegian) or in a hybrid form as monotonal (L) versus bitonal (H+L) (Fretheim 1987; Kristoffersen 1993 for Norwegian; Bailey 1988 for Swedish). Even if I do not deny the correctness of the above observations and admit that my own analysis is not uncontroversial, I still have the following reasons to stick to the symmetric (or equipollent) bitonal and timing analysis of the word accent distinction, i.e. the H+L*/H*+L representation.

A first piece of evidence comes from how we account for the observed pitch patterns of a two word phrase consisting of a combination of an accent II word plus an accent I word as opposed to a corresponding compound word (accent II) with the same, basic rhythmical make-up (stress distribution) as the two word phrase. As has been demonstrated in earlier experimentation both acoustically and perceptually (Bruce 1977), there is a stable pitch difference between these two types of construction which can be described as the presence (in the two word phrase) versus absence (in the compound) of a pitch peak before the second (or final) stress. Phonologically the distinction can be represented as: H*+L . . . H+L* (H$^-$) [two word phrase] versus H*+L . . . L* (H$^-$) [compound]. Examples are *mellan målen* (between the meals) vs. *mellan-målen* (the snacks) and *kassa apparater* (lousy machines) vs. *kassa-apparater* (cash machines). The existence of such a pitch peak has not been denied, but the counter-argument is typically that it must be signalling some kind of word or phrase boundary. However, it can be easily demonstrated by systematically varying location of word boundary and number of inter-stress syllables that the location of this peak is not at the word boundary (other than accidentally) but in the pre-stress syllable, i.e. in the syllable preceding the final stress. This lends support to the interpretation of this pitch peak as part of the accent I gesture.

It should also be emphasized that the distinction between accent I and accent II is maintained not only in focal position but also in non-focal, but still accented positions. While a non-focal accent II is realized with a pitch peak (represented as H*+L), a corresponding non-focal accent I is not just a low level as implied by an L* representation. As a matter of fact it is exactly in a non-focal but accented position that the timing difference between the two word accents is the most apparent, i.e. where we find an early pitch peak for accent I (represented as H+L*).

Furthermore, it is worth pointing out that the absence of this pitch peak, i.e. the H of the H+L* accent I gesture, specifically in a phrase or utterance initial position is not unique. Even if it is clear that this H is missing in this particular context, the same seems to be true of an L boundary tone for an utterance initial accent II word (with initial stress). This would mean that any tone can be skipped in an utterance initial position, either a boundary tone (e.g. %L) or a leading tone of a bitonal pitch accent (e.g. the H of a H+L*). Thus the generalization for Standard Swedish seems to be that leading tones before a starred tone in an utterance initial position will not surface, if there is not enough segmental material (or tone-bearers) before the initial stressed syllable. The same is not true, however, of a phrase or utterance final position. Trailing tones after a final starred tone as for example an H^{-} focal tone and an L% boundary tone will typically be included in the pitch pattern of an utterance final word, even if the consequence of this is tonal crowding. Final lengthening typical of an utterance final position but not for a corresponding initial position may give room for this pitch accent gesture.

Another kind of argumentation which is of a more general nature is that it is more revealing and coherent to describe differences in intonational prominence between Swedish dialects in terms of timing differences of the same basic pitch gesture, as in the Bruce and Gårding model. In the so-called one-peaked dialects (SOUTH and CENTRAL) the word accent distinction may be more transparent as a timing distinction than in the two-peaked dialects (EAST and WEST), where the second peak as the pitch realization of focus could conceal the basic timing distinction. An alternative analysis of the inter-dialectal variation, as has become the traditional account of Norwegian (dialects), is to categorize dialects into H tone and L tone dialects, which refers to the pitch level in the beginning of the stressed syllable of an accent I word. This analysis seems to be less apparent and convincing for a tonal typology of Swedish dialects. For example, the categorization of WEST and CENTRAL dialect types would not be straightforward in this typology. However, the different analyses of Swedish and Norwegian pitch accents do not seem to represent any real differences between the dialects of the two languages but are rather to be thought of as differences in perspective.

A final argument in favour of the timing analysis of the Scandinavian word accent distinction is Einar Haugen's original observation for Norwegian that when studying an Fo contour of an utterance without segmental references it is very difficult if not impossible to identify which is accent I and which is accent II (Haugen and Joos 1952). This will definitely create problems for an automatic recognition of pitch patterns as either accent I or accent II. Haugen's observation speaks in favour of the word accents as consisting of the same basic pitch contour distinguished primarily by their specific timing with respect to the stressed syllables (or feet).

15.3. PITCH REALIZATION OF FOCUS

While in our modelling of (non-focal) accentuation (accent I and accent II) there are several distinct timings of a pitch accent gesture available, the pitch realization of focus represents an either-or feature in an inter-dialectal comparison. Focus is either signalled as a separate pitch gesture added after the word accent gesture proper (EAST, WEST), or as a simultaneous pitch gesture, i.e. as a wider range of the word accent gesture (SOUTH, CENTRAL). It should be added that focal accent is typically accompanied by a concomitant lengthening of the constituent under focus.

The identification of the contribution of focus to the pitch contour as either a wider range of the word accent gesture or as an extra pitch gesture added after the pitch gesture of the word accent itself was a confirmation of the traditional division into single-peaked dialect types (SOUTH, CENTRAL) and double-peaked types (EAST, WEST). The categorization into single-peaked and double-peaked refers to the number of pitch peaks in the citation form of an accent II word (cf. Gårding 1977).

The following table summarizes the primary features of the pitch realization of focus in the four prosodic dialect types of Swedish. These

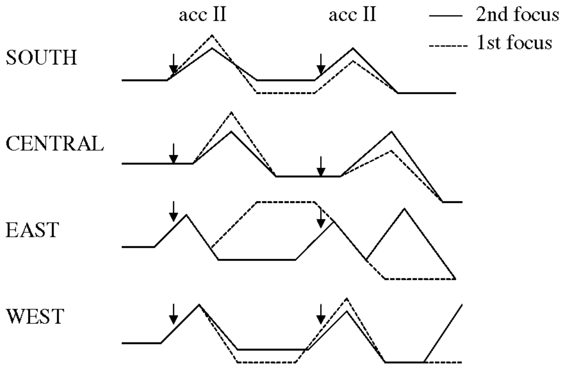


FIGURE 15.5 Focus realization for four prosodic dialect types. Schematic pitch contours of a phrase consisting of two accent II words with focus either on the first (1st focus) or the second (2nd focus) word. The arrow indicates the CV-boundary of stressed syllables.

characteristics hold for non-compound, simplex words. Schematic pitch contours are shown in Figure 15.5. The specific compound pitch patterns will be dealt with in the next section of the paper.

Pitch realization of focus

SOUTH simultaneous gesture having a wider pitch range in focus, with particularly a lower L in non-final positions and a higher H in final positions;

CENTRAL simultaneous gesture having a wider pitch range in focus, with particularly a higher H in all positions;

EAST separate gesture (up to a H^{-}), added relatively early after the word accent gesture, with a potential high concatenation (plateau) in non-final positions between the focal and post-focal accent;

WEST separate gesture (up to a H^{-}), added relatively late after the word accent gesture; in final position realized in the phrase-final syllable, in non-final positions as an updrift between the focal and post-focal accent with the effect of increasing the pitch range of the post-focal accent.

Thus, the wider range of a focal accent in SOUTH and CENTRAL covers the pitch interval between the H turning point and a following L, with either more emphasis on a lower L (SOUTH) or a higher H (CENTRAL). The dialect types with a separate focus gesture seem to display a much clearer category distinction between focal and non-focal accents than the dialect types with a simultaneous focus gesture. This is at least true of EAST (generally) and of WEST in a phrase-final position. In a non-final position the focus gesture of WEST seems to be less categorically distinct. Here it is a widening of the pitch interval between the L after the word accent H of the focal word and the H of the following, post-focal accent. Thus, it is, strangely enough, particularly the pitch range of the post-focal accent that is affected.

A related issue concerns possible, transitional forms between the prototypical prosodic dialect types according to our typology. This question was addressed in outline in the Swedish Prosody project. An example of such a dialect type was found to be a combination of pitch realization of focus as for EAST (type 2A) and of word accent timing as for WEST (type 2B), i.e. somewhat later pitch accent timing than for EAST. This transitional form was termed type 2AB (cf. Gårding *et al.* 1981). The issue of transitional forms will be further explored in the SweDia 2000 project.

It is my impression and a hypothesis for further investigation that there are more instances of equal weight among accented words within a phrase in SOUTH and CENTRAL. For SOUTH we can compare with Danish, a neighbouring Scandinavian dialect, where there is no particular focal accent (sentence accent) among accented words within a phrase, except for special cases with emphasis for contrast (Grønnum 1992).

The above taxonomy gives us an account of the paradigmatic relations of focal accentuation in an inter-dialectal comparison. But it should be pointed out that phonetic focus is primarily a syntagmatic relation between successive, prominent constituents within a prosodic domain (phrase). Therefore, it should be emphasized that the pitch realization of focus is not only a tone or a pitch gesture added to a focal constituent but as much its relation to the surrounding non-focal, but still prominent constituents. The syntagmatic nature of phonetic focus has been given proper emphasis in works by Strangert and Heldner (1997). Thus for example phrase-internal downstepping pitch patterns may belong to the signalling of focus. This seems to occur in dialect types where the particular focus signal constitutes a separate pitch gesture. For EAST Swedish downstepping within a phrase is a feature of successive post-focal accents but is not present in pre-focal position. In dialect types where there is no such added focus gesture, like in SOUTH, deaccentuation (lack of intonational prominence) after an early focus of a phrase has a corresponding weighting effect. It is interesting to note that the pitch patterns of focus signalling in SOUTH (no separate focus gesture; post-focal deaccentuation) are generally more similar to those of most European languages, e.g. English, French, Greek, than in EAST (extra, separate focus gesture; post-focal downstepping).

15.4. PITCH PATTERNS OF COMPOUNDS

Compounding is a very productive process of word formation in Swedish. In our opinion, compound words are particularly revealing for intonational differences among prosodic dialect types (cf. Bruce and Gårding 1978; Gårding and Bruce 1981). Thus, another decision in the dialect typology is whether the pitch patterns of compounds are distinct from those of simplex words. There is a basic distinction between simplex and compound words in terms of their rhythmical structure. Whereas simplex words are characterized by having one true word stress, a compound word has got (at least) two word stresses: typically a primary (main) stress on the first element of the compound and a secondary stress on the last element of the word. In connection with the primary stress of the compound we find the regular pitch pattern of word accent. In most dialects a compound takes accent II, but in some dialects (SOUTH) it can be either accent I or accent II depending on the phonological/morphological structure of the word. The main regularity is instead that the accent of the first element of the compound (as a separate word) will also be the resulting accent of the compound, but the subregularities are fairly complex (cf. Bruce 1973).

For the pitch patterns of compounds the status of the secondary stress is the main difference between the dialect types in our typology. While there appears to be no distinction in terms of the pitch patterns of compounds as opposed to those of simplex words in WEST and SOUTH, the secondary stress can be shown to be a relevant synchronization point of pitch gestures for EAST and CENTRAL. For the illustration of the varying intonational structure of compounds I will make a distinction between what can be called 'short' and 'long' compounds. The critical factor here is the rhythmical structure, specifically the distance between the two stresses of a compound, namely the primary and the secondary stress. While in a short the two stresses occur in immediate succession (clash), in a long compound a number of unstressed syllables tend to separate the primary and secondary stresses (see Figure 15.6).

In EAST Swedish secondary stress is a critical point of synchronization of the focal accent rise (LH). Whereas in a simplex accent II word the focal accent rise (LH) is executed immediately after the word accent fall (H*+L), the secondary stress is instead the trigger of the focal rise in an (accent II)

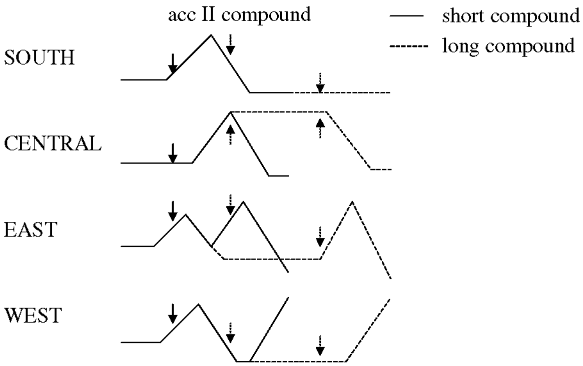


FIGURE 15.6 Pitch patterns of compounds for the four dialect types. Schematic pitch contours of short/long accent II compounds in focal and phrase-final position. The first arrow marks the CV-boundary of the primary stress, and the second/third arrow marks the CV-boundary of the secondary stress of the word.

compound. In more complex compounds containing more than two stresses, the secondary stress is the final stressed syllable of the word. This means that the pitch patterns of compounds and simplex words are clearly distinct, particularly when the secondary stress is well separated from the primary stress by means of a number of unstressed syllables or other intermediate stresses. In a non-focal position there is no intonational difference between a compound and a corresponding simplex word, but only a rhythmical difference.

It is interesting to note that Stavanger Norwegian (cf. Fintoft 1970) which appears to have the same basic pitch realization of the word accents as EAST (Stockholm) Swedish, does not seem to show the distinction in pitch synchronization between non-compound and compound words (pers. comm. Tomas Riad). The pitch rise up to an H comes immediately after the word accent fall (H+L) even in compound words. The same is also true of SOUTH Swedish speakers trying to acquire a standard EAST Swedish accent. Typically, they do not respect the secondary stress of a compound as a pitch trigger of the rise to a focal accent H^{-} but tend to apply an early focal accent rise.

In CENTRAL Swedish secondary stress in compounds is also a relevant point of pitch synchronization. Unlike EAST this has nothing to do with whether it is a focal or non-focal position. The primary stress is characterized by the regular accent II gesture (L*+H). While in a simplex word after the accent II rise (L*+H) there will be an almost immediate decrease in pitch to an L, this is not the case in a compound word. After the initial rise (L*+H) pitch will stay high (and can form a plateau) in a compound until the secondary stress, which is the trigger of a pitch fall (HL). In this dialect type the difference between focal and non-focal position is a matter of relative pitch range, as has been described above.

Also dialects of FAR EAST, i.e. Finland Swedish, which are characterized by having no word accent (accent I/accent II) distinction, behave like WEST and SOUTH in having no pitch distinction between simplex and compound words.

A possible extension of our prosodic dialect typology is to recognize NORTH Swedish as a prosodic dialect type distinct from EAST Swedish. These dialects—NORTH and EAST—are prosodically similar in several respects, but the stress distribution in compounds may be different, particularly in the far NORTH Swedish dialects. Compounds other than those with an internal stress clash display a rhythmical switch as it were. While in most Swedish dialects compound words are characterized by an early primary stress and a late secondary stress, in far NORTH Swedish stress distribution in many compounds is the other way around: an early secondary stress and a late primary stress, or even only one (primary) stress located in the final element of the compound. While the initial stress has no relevant pitch gesture (and thus no word accent), only the final, primary stress is characterized by a pitch accent, which is always accent I (H+L*). This final stress is then the relevant synchronization point for the word accent gesture (accent I) as well as for the succeeding focal accent gesture (up to a H^{-}), when the word is in focus.

15.5. SYNTHETIC SIMULATION

To illustrate the importance of intonation for dialect identity and to test our prosodic dialect typology we have made an attempt to simulate the different

*(de’ e’ en syn**te**tisk dia**le**kt som **da**tatek**ni**ken **fi**xat)*
[deɛn sʏnˈˈteːtɪsk dɪaˈlɛkt sɔm ˈˈdàːtatɛkˌniːkən ˈfìksat]
(It’s a synthetic dialect that was fixed by computer technique)

Simulation of six Swedish dialect types:
SOUTH (Syd), CENTRAL (Dal), EAST (Sve), WEST (Göt), NORTH (Nor), FAR EAST (Fin)

FIGURE 15.7 Synthetic test utterance in Swedish in orthography (stressed syllables in boldface) and phonetic transcription (IPA). Simulation of six Swedish dialect types.

dialect types using resynthesis of intonation. A phonetically balanced and relatively dialect neutral test utterance was chosen for the resynthesis, where focus location is denoted by capital letters: *de' e' en SYNTETISK dialekt som DATATEKNIKEN fixat* (It's a synthetic dialect that was fixed by computer technique). The informant who recorded the test utterance is a female speaker of WEST Swedish. Her instruction was to speak as naturally as possible but without intonational variation (Figure 15.7).

Variables in the intonation model were different timings of the word accent gesture (several points on a scale early–late), simultaneous/separate focus gesture, and same/different pitch pattern for simplex and compound words. The following diagram shows the parameter values that have been chosen for the simulation of six different prosodic dialect types of Swedish (see also Figure 15.8).

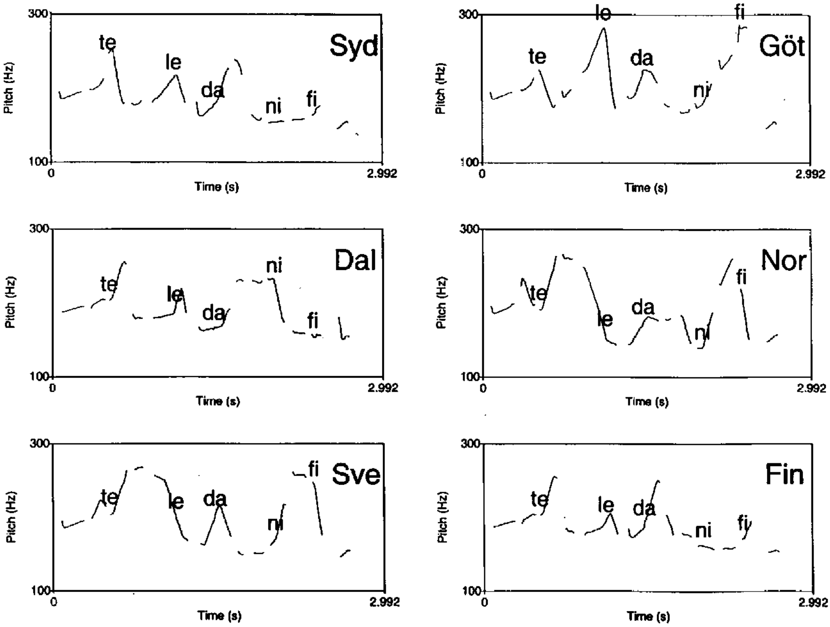


FIGURE 15.8 Simulation of six Swedish dialect types: *SOUTH (Syd), CENTRAL (Dal), EAST (Sve), WEST (Göt), NORTH (Nor), FAR EAST (Fin)*. Pitch contours of the resynthesized utterance with indication of location of successive stressed syllables. (Sound files are included; from Figure 15-8a.wav. to Figure 15-8f.wav. The sound file of the original sentence is included in Figure 15-8.wav.)

	accent	*focus*	*compound*
SOUTH	late timing	simultaneous	not distinct
CENTRAL	very late timing	simultaneous	sec. stress relevant
EAST	early timing	separate	sec. stress relevant
WEST	central timing	separate	not distinct
NORTH	early timing	separate	final stress
FAR EAST	late timing (no acc I1acc II distinction, timing like acc **II in** SOUTH)	simultaneous	not distinct

No formal testing of the synthetic versions of the different dialect types has been undertaken so far, but the reactions among those exposed to the synthetic dialects show rather unanimously that the simulation of the first half of the dialects is successful, while for the other halfthe simulation is still prosodically incomplete. Our interpretation of this is that we have at least a partial knowledge of the prosodic variation among Swedish dialects, but extended research is needed for a more complete understanding of this variation.

15.6. FUTURE WORK

In the SweDia 2000 project our starting point was the intonational modelling of prosodic dialect types within the Swedish Prosody project (cf. Bruce and Garding 1978) and later work **in** the area (cf. Riad 1998; Engstrand 1995,1997). By using new data collected **in** the SweDia 2000 project we hope to be able to develop the prosodic dialect typology for Swedish dialects. An example of a hypothesis that was tested on the SweDia 2000 data base is the idea that prosody, **in** particular accentuation and intonation, varies little within a major regional dialect area, while other phonetic features like vowel quality are more variable locally within such a dialect area.

REFERENCES

ARVANITI, A., LADD, D. R., and MENNEN, 1. (1998), 'Stability of Tonal Alignment: The Case of Greek Prenuclear Accents', *Journal ofPhonetics,* 26: 3-25.

BAILEY, 1. (1988), 'Representing Pitch Accent in Swedish', *PERIL US* VIII (Stockholm: University of Stockholm, Department of Linguistics), 153-83.

BRUCE, G. (1973), 'Tonal Accent Rules for Compound Words in the Malmö Dialect', *Working Papers* 7 (Lund: Lund University, Phonetics Laboratory), 1-35.

BRUCE, G. (1977), *Swedish Word Accents ın Sentence Perspective* (Lund, Sweden: Gleerup).

— — (1987), 'How Floating is Focal Accent?', in K. Gregersen and H. Basb0ll (eds.), *Nordic Prosody* IV (Odense: Odense University Press), 41-9.

— — (forthcoming), 'Tonal Variation in Swedish', to appear in Shigeki Kaji (ed.), *Proceedings of the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Historical Development, Phonetics of Tone and Descriptive Studies* (Tokyo: ILCAA, Tokyo University of Foreign Studies), December 2002.

— —, ELERT, C.-C., ENGSTRAND, O., ERIKSSON, A., and WRETLING, P. (1999), 'Database Tools for a Prosodic Analysis of the Swedish Dialects', in *Proceedings Fonetik* 99 (Gbteborg: Gbteborg University, Department of Linguistics), 37-40.

— —, and GARDING, E. (1978), 'A Prosodic Typology for Swedish Dialects', in E. Garding, G. Bruce, and R. Bannert (eds.), *Nordic Prosody* (Lund, Sweden: Lund University, Department of Linguistics), 219-28.

D'IMPERIO, M., and HousE, D. (1997), 'Perception of Questions and Statements in Neapolitan Italian', in *Proceedings EUROSPEECH* '97 (Rhodes, Greece), 1: 251-4.

ELERT, C.-C. (1964), *Phonologic Studies of Quantity in Swedish* (Stockholm: Almqvist & Wiksell).

ENGSTRAND, O. (1995), 'Phonetic Interpretation of the Word Accent Contrast in Swedish', *Phonetica,* 52: 171-9.

— — (1997), 'Phonetic Interpretation of the Word Accent Contrast in Swedish: Evidence from Spontaneous Speech', *Phonetica,* 54: 61-75.

— —, BANNERT, R., BRUCE, G., ELERT, C.-C., and ERIKSSON, A. (1997), 'Phonetics and Phonology of Swedish Dialects Around the Year 2000: A research Plan', *Reports from the Department of Phonetics* (Umeå: Umeå University), 97-100.

FINTOFT, K. (1970), *Acoustical Analysis and Perception of Tonemes in some Norwegian Dialects* (Oslo: Universitetsforlaget).

FRETHEIM, T. (1987), 'Phonetically Low Tone-Phonologically High Tone, and Vice Versa', *Nordic Journal of Linguistics,* 10(1): 35-58.

GÅRDING, E. (1977), *The Scandinavian Word Accents* (Lund, Sweden: Gleerup).

— — (1982), 'Swedish Prosody', *Phonetica,* 39: 288-301.

— —, and BRUCE, G. (1981), 'A Presentation of the Lund Model for Swedish Intonation', *Working Papers* (Lund: Lund University, Department of Linguistics), 21: 69-75·

— —, BRUCE, G., and WILLSTEDT, U. (1981), 'Transitional Forms and their Position in a Prosodic Typology for Swedish Dialects', *Working Papers* (Lund: Lund University, Department of Linguistics), 21: 77-87.

— —, and LINDBLAD, P. (1973), 'Constancy and Variation in Swedish Word Accent Patterns', *Working Papers* (Lund: Lund University, Phonetics Laboratory), 7: 36-110.

GR0NNUM, N. (1992), *The Groundworks of Danish Intonation* (Copenhagen: Museum Tusculanum Press).

HAUGEN, E., and *Joos,* M. (1952), 'Tone and Intonation in East Norwegian', *Acta Philologica Scandinavica,* 22: 41-64.

HOUSE, D. (1990), *Tonal Perception in Speech* (Lund, Sweden: Lund University Press).

KOHLER, K. (1987), 'Categorical Pitch Perception', in Ü. Viks (ed.) *Proceedings ICPhS* 11 (Tallinn: Academy of Sciences of the Estonian SSR), 331-3.

KRISTOFFERSEN, G. (1993), 'An Autosegmental Analysis of East Norwegian Pitch Accent', in B. Granstrom and 1. Nord (eds.), *Nordic Prosody* VI (Stockholm: Almqvist & Wiksell), 109-22.

LADD, D. R. (1983), 'Phonological Features of Intonational Peaks', *Language, 59:* 721-59·

— — (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

MEYER, E. A. (1937), *Die Intonation im Schwedischen: Die Sveamundarten* (Studies Scand. Philo!' 10) (Stockholm: University of Stockholm).

— — (1954), *Die Intonation im Schwedischen: Die norrlandischen Mundarten* (Studies Scand. Philo!' 11) (Stockholm: University of Stockholm).

OHMAN, S. (1967), 'Word and Sentence Intonation: A Quantitative Model', *Speech Transmission Laboratory, Quarterly Progress and Status Report,* 2-3 (KTH, Stockholm), 20-54.

PIERREHUMBERT, J., and STEELE, S. (1989), 'Categories of Tonal Alignment in English', *Phonetica,* 46: 181-96.

RIAD, T. (1998), 'Towards a Scandinavian Accent Typology', in W. Kehrein and R. Wiese (eds.), *Phonology and Morphology ofthe Germanic Languages* (Tubingen: Max Niemeyer), 77-109.

STRANGERT, E., and HELDNER, M. (1997), 'The Contribution of Pitch Movements to Perceived Focus', *PHONUM* 4 (Umeå: Umeå University, Department of Phonetics), 109-12.

WITHGOTT, M., and HALVORSEN, P.-K. (1988), 'Phonetic and Phonological Considerations Bearing on Representation of East Norwegian Accent', in H. van der Hulst and N. Smith (eds.), *Autosegmental Studies on Pitch Accent* (Dordrecht: Foris), 279-94.

16

Prosodic Typology

Sun-Ah fun

16.1. INTRODUCTION

Studies on prosodic typology are in general rare probably because prosodic features are not easy to define and categorize, and also because prosodic features oflanguages have been described, if at all, with different assumptions and within different frameworks. Finding similarities and differences of prosodic features across languages would make sense only if these languages were described in the same framework in terms of the same prosodic categories. Comparisons of a prosodic system based on phonetic descriptions would have limitations because the similarities shown in the surface realization do not guarantee the same underlying distinctive prosodic features or structures. Since the advent of the phonological model of intonation in the 1980s, especially, the Autosegmental-Metrical (AM) model of intonational phonology, prosody has been described in terms of a prosodic structure and distinctive tonal categories (e.g. Pierrehumbert 1980; Gussenhoven 1984; Liberman and Pierrehumbert 1984; Beckman and Pierrehumbert 1986; Pierrehumbert and Hirschberg 1990; Ladd 1992,1996). This has made it much easier to compare the prosody of languages categorically and closer to the goal of establishing prosodic typology (e.g. Ladd 2001)

In the AM model, prosody is described in two aspects: the prosodic structure of an utterance and the prominence relations within the structure (d. Beckman 1996; Ladd 1996; Shattuck-Hufnagel and Turk 1996; Fougeron 1998; Jun *2oo3a).* A prosodic structure is a hierarchical organization of prosodic units from the smallest prosodic unit (Mora or Syllable) to the largest (Intonation Phrase or Utterance). Within a phrase, some words are more prominent than others; and

I would like to thank Bob Ladd for his valuable comments on an earlier version. I would also like to thank those who have provided me with their papers and their expertise. They are in alphabetical order: Gösta Bruce, Dana Chahal, Mariapaola D'Imperio, Gorka Elordieta, Jennifer Fitzgerald, Janice Fon, Sonia Frota, Svetlana Godjevac, Matt Gordon, Larry Hyman, Kikuo Maekawa, Aniruddh Patel, Shu-hui Peng, Pilar Prieto, Stefanie Shattuck-Hufuagel, and Jennifer Venditti.

within a word, some syllables are more prominent than others. A prosodic structure and prominence relations are realized by suprasegmental features such as pitch, duration, and/or amplitude as well as segmental properties such as the realization of consonants and vowels. Furthermore, the prosodic property of an utterance is a combination of prosody at the word level and prosody at the phrase level. Postlexical prosody is constrained by the lexical prosody, and postlexical prosodic information contains information about the lexical prosody.

One of the most well-known properties of prosodic typology at the lexical level is word prosody: whether a lexical item has tone, stress, or lexical pitch accent (Trubetzkoy 1939; Beckman 1986; Ladd 1996; Fox 2000 and references therein). Under this typology, languages have been categorized as tone languages such as Mandarin and Hausa, as stress languages such as English and German, or as lexical pitch-accent languages such as Japanese and Basque. However, as shown in the previous chapters, a language can be specified with more than one such lexical feature or specified with none of them (see also Remijsen 2001). For example, a tone language can have stress, as in Mandarin, or not, as in Cantonese. Languages known to have stress can also have lexical pitch accent, as in Chickasaw, or not, as in English and other West Germanic languages. Languages known to have lexical pitch accent can have stress, as in Serbo-Croatian and Swedish, or not, as in Japanese. Finally, non-tonal and non-lexical-stress languages can have lexical pitch accent, as in Tokyo Japanese, or not, as in Seoul Korean. In sum, whether a language has stress or not is independent of whether a language has tone or lexical pitch accent specification.

Furthermore, all such lexical features interact with postlexical prosody, especially intonation, and intonational features are not directly predictable from the lexical prosodic features. That is, a pitch event in a syllable can provide postlexical information such as postlexical pitch accent (or sentence stress), a phrasal tone, or a boundary tone. Postlexical pitch accent can be from lexical pitch accent (e.g. Japanese) or from stress, either lexical (e.g. English and most other West Germanic languages) or postlexical (e.g. French). A phrasal tone, which marks a prosodic unit such as a Prosodic Word or an Accentual Phrase, can be found in languages with lexical pitch accent (e.g. Japanese, Serbo-Croatian), stress (e.g. Chickasaw, Farsi) or with no lexical specification (e.g. Korean). The postlexical pitch event can also differ in its function. It can mark prominence (e.g. pitch accent in English and other Germanic languages), demarcate a boundary of a prosodic unit (e.g. the accentual phrase in Japanese and Korean, the intonation phrase boundary tone in most languages), or both (e.g. demarcating pitch accent in French; Jun and Fougeron 1995,2000,2002). Therefore, in order to study prosodic typology, we need to examine postlexical prosodic features as well as the lexical prosodic features.

Another well-known property of prosodic typology is the rhythm or timing unit ofa language. Languages have been categorized as mora-timed as in Japanese, syllable-timed as in Spanish, or stress-timed as in English (e.g. Pike 1945; Bloch 1950; Abercrombie 1967; Lehiste 1970). Though studies have shown that isochrony between syllables or between stress intervals does not have acoustic correlates and the division between syllable-timed and stress-timed is not straightforward (e.g. Lehiste 1977; Cutler 1980; Nakatani *et al.* 1981; Beckman 1982; Roach 1982; Dauer 1983,1987; Cooper and Eady 1986), recent studies on the variation in vowel duration show that syllable-timed languages show less contrast between two consecutive vowel durations than stress-timed languages (Ramus *et al.* 1999; Low *et al.* 2000; Grabe and Low 2002; Ramus 2002).[1]

Even though our perception of the prosody of a language is influenced by the rhythm unit smaller than a word (Mora for mora-timed, Syllable for syllable-timed, or Foot for stress-timed), it is also influenced by the prosodic grouping at or above the Word such as a Prosodic Word, an Accentual Phrase, or an Intonational Phrase. A prosodic unit above the Word also has a rhythmic nature. Each unit often begins or ends with a prominent syllable or word, either due to pitch accent, a boundary tone, or both. A small prosodic unit such as an Accentual Phrase often has only one pitch accent, and is either followed by unaccented words (e.g. Japanese) or preceded by unaccented words (e.g. French); and similarly, an Intonation Phrase in stress accent languages often ends with a nuclear pitch accent, the most prominent pitch accent in the phrase. Furthermore, the edge of a prosodic unit above the Word tends to occur at a regular interval, and each prosodic unit avoids being too short (like stress clash) or too long (like stress lapse) unless it is influenced by other factors such as focus. For example, in Korean the size of the Accentual Phrase is three to five syllables on average, and an Accentual Phrase shorter than three syllables or longer than seven syllables is very rare (Korean Telecom 1996; Kim *et al.* 1997; Jun and Fougeron 2000).

As we cannot predict the postlexical prosody of pitch (e.g. intonational pitch accent, phrase accent) based on the lexical prosody (i.e., tone, stress, lexical pitch accent) of a language, the prosodic units above the Word are not predictable from the timing unit ofthe language. A language can have an Accentual Phrase, a small prosodic unit above the Word, whether it is mora-timed (e.g. Japanese), syllable-timed (e.g. French), or stress-timed (e.g. Chickasaw).

[1] However, this method does not tell what rhythmic unit a language has, e.g. syllable, foot. It only suggests a rhythmic unit of a language indirectly through the variability of vowel duration. A language with little contrast in vowel duration would be perceived as syllable-timed, but it is not clear if a language with more contrast in vowel duration is always perceived as stress-timed. More languages need to be examined to see if there is a strong correlation between the rhythmic unit and the degree of contrast in vowel duration.

As seen in previous chapters, prosodic units above the Word are often defined by intonation, but the analyses of tone languages such as Cantonese and Mandarin suggest that these languages do not have a tonal event consistently marking a prosodic unit. In Mandarin, an utterance is not always marked by a boundary tone, and there is no tonal event marking a prosodic unit within an utterance. But, the inventory of break indices suggests that speakers of this language perceive at least one level of prosodic grouping between the Word (syllable) and the Utterance (e.g. minor group, major group). This indicates that whether a prosodic grouping larger than the Word is marked by intonation or not is influenced by the functional load of pitch in the language (d. Jun 1998). In sum, to study prosodic typology, prosodic units defined by the degree of juncture, i.e., break index in ToBI (Tones and Break Indices), should be included along with prosodic units defined by intonation.

In this chapter, I will attempt to provide prosodic typology by comparing the prosodic features oflanguages described in this book and several other languages which have been described in the AM model ofIntonational phonology. Prosodic features which are common to these languages as well as those specific to a certain language will be discussed. The prosodic features of eleven languages whose ToBI transcription systems are described in this book (see Chapter 2 for detailed descriptions of ToBI) will be discussed in more detail. Finally, I will discuss the flexibility and the extension of the ToBI system: how the ToBI system has been extended to incorporate different prosodic systems while maintaining its integrity.

The organization of this chapter is as follows. Section 16.2 presents a summary of the prosodic systems of eleven languages based on the ToBI models described in this book, and Section 16.3 proposes a model of prosodic typology based on aspects of prominence and rhythmic/prosodic structure. Section 16.4 presents the flexibility of and other issues regarding the ToBI transcription system, and Section 16.5 concludes the chapter.

16.2. SUMMARY OF ELEVEN ToBI SYSTEMS

Table 16.1 shows a summary of the eleven ToBI systems described in the previous chapters, focusing on the tones and break indices. It includes the ToBI systems of Mainstream American English (MAE_ToBI), standard German (GToBI), Athens Greek (GRToBI), Neapolitan Italian (IToBI),[2]

[2] The chapter on Italian ToB! (Chapter 13) describes the tonal patterns and prosodic structure oHour dialects (three Southern dialects, Neapolitan, Bari, and Palermo, and a central one, Florentine). Neapolitan is chosen arbitrarily. It is claimed that the other dialects use the same tonal shapes (falling, rising, high, and low) but that the tonal inventories are not identical due to different tone-text alignments.

TABLE 16.1 Summary of the eleven ToBI systems introduced in the previous chapters

Lang.	Types of tiers—extra only	Types of break indices (BI)	Types of tones on the tones tier	Prosodic units
English		0, 1, 2, 3, 4	L*, H*, L+H*, L*+H, H+!H*	
			L -, H-	ip
			L%, H%	IP
			! (for H pitch accent), <, >	
German		3, 4,	L*, H*, L+H*, L*+H, H+!H*, H+L*	
		2r (rhythm mismatch),	L-, H-, !H-	ip
			%, L%, H%, ^H%	IP
		2t (tone mismatch)	!, ^ (both for H pitch accent), <, >	
Greek	Prosodic word (=phonetic transcription)	0, 1, 2 (ip),	L*, H*, L+H*, L*+H, H*+L	
		3 (IP),	L-, H-, !H-	ip
		s (sandhi),	L%, H%, !H%	IP
		m (mismatch)	! (for *), <, >, w (for L* undershoot)	
Italian (Neapolitan)		0, 1, 2, 3, 4	L*, H*, L+H*	
			L*n, L+H*n, L*+Hn, H+L*n	ip
			H(*)L-	ip
			L%	IP
			! (for *), n (for nuclear pitch accent)	
Serbo-Croatian	Glosses	0, 1, 2 (IP),	L*+H, H*+L	Wd
		m (mismatch)	%L, %H	Wd
			Ø-, LH-, L%, H%, HL%	IP
			>, # (pitch range rising)	
Japanese	Finality	0, 1, 2 (AP),	H*+L	AP
		3 (IP),	H-, L%, %L	AP
		finality,	H%, LH%, HL%	IP
		m (mismatch)	<, >, w (for L% or %L undershoot)	

Table 16.1 (*Continued*)

Lang.	Types of tiers—extra only	Types of break indices (BI)	Types of tones on the tones tier	Prosodic units
Korean	Phonol tone,	0, 1, 2 (AP),	L, H, +H, L+, Ha, La, LHa	AP
	Phonetic tone	3 (IP),	L%, H%, LH%, HL%, LHL%,	IP
		m (mismatch)	HLH%, LHLH%, HLHL%, LHLHL%	
			<, >	
Mandarin	Romanzi,	0, 1,		
	Syll,[a]	2 (minor grp),	L%, H%, %reset, %q-raise, %e-prom, %compress	Breath group
	Stress,	3 (major grp),		
	Sandhi,	4 (breath grp),		
	Code	5 (prosodic grp)[b]		
Cantonese	Syllable, Foot[c]	0, 1, 2 (IP)	lexical tones (55, 33, 22, 335, 223, 221, 553)	Wd
			L%, H%, H:%, HL%, %, -%, %fi (frame initial boundary)	IP
Chickasaw	Phonetic transcription	0,	H$^\lambda$	Wd
		1 or 2 (AP),	LHHL, HL, LL, LHH	AP
		3 (IP),	H*, !H*	IP
		m (mismatch)	L%, H%, HL% <	IP
BGW	Gloss	0, 1, 2,	H*, H*<, ^H*, L+H*	
	(Syll)	3 (IP),	Lp	PhP
		4 (Utt)	L%, H%, LH%, %L, %H	IP
			! (for * tones)	

[a] Unlike Cantonese ToBI, four lexical tones (55, 51, 32, 124) are labelled in the Romanzi (Romanization) tier in Mandarin.

[b] The authors note that a large percentage of inter-transcriber disagreements involved confusion between BI 4 and BI 5. They suggest that these two labels may be collapsed later.

[c] In Cantonese, the Syllable tier tags an alphabetic transliteration for every syllable. It is similar to Mandarin ToBI's Romanzi tier (Mandarin ToBI's Syllable tier tags phonological syllables). The Foot tier tags syllable fusion and the domain of emphasis. It does not function as a prosodic unit and differs from the foot in stress languages. Cantonese ToBI recommends three more tiers for some sites: Phones, Sociolinguistic variables, and Code.

Serbo-Croatian (SCToBI), Tokyo Japanese (J_ToBI), Seoul Korean (K-ToBI), Mandarin (M_ToBI), Cantonese (CToBI), Chickasaw, a Western Muskogean American Indian language (Ch-ToBI), and Bininj Gun-wok, an indigenous Australian language, also known as Mayali (BGW ToBI).

To compare the prosodic systems of these languages, their ToBI systems are compared in four categories: types of tiers, types of break indices, types of tones on the Tones tier, and types of prosodic units defined by intonation (not by juncture which is represented by break indices).

For the Types of Tiers, only the tiers other than the four original tiers (Words, Tones, Break Indices, Miscellaneous) proposed in American English ToBI are provided. This is because all the ToBI systems described in this book have these four tiers and only some have additional tiers, reflecting the different prosodic systems of the languages.

For the types of break indices, the numerical number is given when the meaning of the index is the same as or similar to that of the English break index (i.e., 0 for a weakened word boundary, 1 for a phrase-medial word boundary, 3 for a minor phrase boundary such as an Intermediate Phrase, 4 for a major phrase boundary such as an Intonation Phrase, and 2 for mismatch). When the meaning of any break index is different from this default meaning, the meaning (i.e., the name of a prosodic unit marked by the break index) is given in parentheses right after the break index (e.g. 2(IP) in Serbo-Croatian means the break index 2 marks the end of an Intonation Phrase). Diacritics for the break indices such as 'p' (disfluent pause), '–' (ambiguity between two indices), and '?' (uncertainty) are not included because most ToBI systems include these labels and these labels have the same meaning. However, diacritics specific to a certain ToBI system, such as 's' in Greek, are included with the interpretation in parentheses (e.g. s (sandhi)).

For the types of tones on the Tones tier, tones in each row have the same function and belong to the same prosodic unit. That is, they are either lexical tones (e.g. Chinese tone and Chickasaw lexical pitch accent), the head of a prosodic unit such as pitch accent (marked by *), or the boundary tone marking the edge of a prosodic unit such as an Accentual Phrase (AP), an Intermediate Phrase (ip), or an Intonation Phrase (IP). For Mandarin and Cantonese, tones for marking pitch range information (e.g. %reset, %q-raise) are grouped together with the boundary tone of the highest prosodic group. The name of the prosodic unit corresponding to each tone type is given in the last column, Prosodic Units, on the same row. The prosodic unit corresponding to the postlexical pitch accent is not given because the domain of the postlexical pitch accent is not known (cf. see Beckman and Edwards 1990 for a discussion of English pitch accent domain). Diacritics were added to the tone type if they are used as a part of a distinctive tone type as in English H+!H* or German ^H%. However, if diacritics were used with more than one tone type, the symbol is given in the last row within the tones column for each language (e.g. '! (for H pitch accent)' or '! (for *)', meaning downstep can happen to all H pitch accent tones).

The summary table shows that all languages have prosodic units above the word. Except for the two tone languages (Mandarin and Cantonese), all have two prosodic units marked by intonation. Though these languages differ in the number of prosodic units above the Word, ranging from one (Serbo-Croatian and Cantonese) to four (Mandarin), most of them have two prosodic units above the Word: the largest phrase, often called Intonation Phrase, and a smaller phrase, called by various names such as Intermediate Phrase, Accentual Phrase, or Phonological Phrase. The prosodic units higher than the Word are often identified by intonation as well as by the degree of juncture, as implied by the fact that a break index and a boundary (or phrasal) tone mark the same prosodic unit. However, this mapping is not perfect in natural speech as indicated by the use of a break index (BI 2 or a diacritic 'm') for mismatch cases in most ToBI analyses. Currently, the two ToBI analyses of tone languages do not include any mismatch labels probably because tonally defined prosodic units smaller than an IP or Breath Group do not exist or have not been identified yet.

The ToBI analyses of languages also show that languages differ in the types of tones (lexical tone, pitch accent, or boundary tone) and tonal inventories they have. The tonal types can tell us whether a language is a tone language, an accent language, or whether it has no lexical specification of prosody. The current version of ToBI systems proposed for contour tone languages (Mandarin, Ch. 9, and Cantonese, Ch. 10) shows that they do not have pitch accent (a * tone) and that they have smaller 'intonational' tone inventories compared to other languages.[3] They only have a few boundary tones at the edge of the largest prosodic unit and have some labels marking pitch range specifications. The language with no lexical specification of prosody (Korean) does not have a * tone, either, but has the largest variety of boundary tones. The types of tones, however, cannot distinguish stress-accent languages from lexical pitch-accent languages. This is because the AM model does not specify whether pitch accent is a lexical property or a postlexical property. A major difference between these languages, based on the summary table, is in the number of tonal inventories. Stress-accent languages have multiple types of pitch accent while lexical pitch-accent languages have only one or two types of pitch accent. Among the languages which include pitch accent, H* is the

[3] It could be possible to propose a postlexical pitch accent in Mandarin when a syllable receives sentence level stress. A ToBI system of African tone languages would be needed to get a better picture of the prosody of tone languages in general as well as the prosodic properties specific to Chinese-like contour tone languages.

most common and L* is the least common. For all these languages, high or rising tones (a sequence of L and H) are more common than falling tones, and single boundary tones (i.e., L%, H%) are more common than multiple boundary tones (e.g. LH%, LHL%).

For tiers, three types of tiers have been added in addition to the original four tiers. A first type is to represent language-specific prosodic features, a second type is to help labellers who are not native speakers of the language, and a third type is to label information needed to investigate the interaction between prosody and sub-areas of linguistics such as the role of prosody in syntax-phonology mapping, discourse structure, and dialect differences. The first type includes the syllable-related tiers such as the Syllable and the Stress tiers in a tone language (where each syllable carries a tone, and the degree of stress affects the realization of the underlying tone) and the Phonetic Tone tier in Korean (where some of the phrasal tones show variations in tonal categories, categorically affecting the tonal shape of the phrase, but have no distinctive function). The second type includes the English Gloss tier in some of the non-English languages (e.g. Serbo-Croatian, BGW). The third type includes the Phonetic Transcription tier where allophones are specified to study the domain of a segment sandhi (e.g. Greek, Chickasaw), the Code tier for a dialect specification (e.g. Mandarin) to study code switching and dialect typology, and a Finality tier for marking discourse finality (e.g. Japanese) to study the relationship between prosody and discourse structure.

Finally, the summary table shows that the Tones tier uses diacritics to cover both distinctive tonal targets and the phonetic realizations of the tonal targets. The diacritics for the local pitch range adjustments (e.g. !, ^, #) are distinctive in some languages (e.g. H+!H* in English and German, ^H* in BGW) but not in others (e.g. !H* in Greek and Chickasaw). Similarly, the diacritics for timing in the tone-text alignment (e.g. '<' for a peak delay and '>' for an early peak) are used to mark a distinctive tonal category in Bininj Gun-wok (H*<), but are used to mark a phonetic event in all others. This suggests that the function of diacritics in the ToBI system is specific to its variety. This issue will be further discussed in Section 16.4.

Though not represented in the summary table, languages also differ in their tone-text association and their alignment. For example, the H- phrase accent in Greek interrogatives is associated with a stressed syllable, if available, but not in English. The L*+H pitch accent in Greek is realized as a rising tone with the fo minimum before the stressed syllable and the fo maximum after the stressed syllable (Arvaniti *et al.* 1998), but the same pitch accent is realized in English as a rising tone with the fo minimum *during* the stressed syllable and fo maximum after the stressed syllable (Pierrehumbert and Steel

1989; Hirschberg and Ward 1992; Beckman and Ayers-Elam 1997). Similarly, the realization of the same prosodic unit also differs across languages. For example, a falling tone (HL%) can mark the end of an Intonation Phrase in Japanese and Korean, but not in English, German, or Greek. The end of an Intonation Phrase is significantly lengthened in English and most other languages, but it is not always lengthened in Japanese. A High tone can mark the end of an Accentual Phrase in Korean, but not in Japanese.

Furthermore, languages, and even the dialects within the same language, can differ in their mapping between a tone and the meaning. Though there are cases where a tone-meaning mapping is fairly consistant across languages or dialects (e.g. a low boundary tone for a statement, a high boundary tone for a yes-no question), the mapping is specific to a language and a dialect. A basic meaning of a certain tone type (e.g. boundary tones, focus tone) is given in each ToBI system, but given that the relationship between a tone and the meaning is not one-to-one, but many-to-many even in a single dialect, it would be premature to include tone-meaning relations in prosodic typology. In order to find out tone-meaning mappings across languages, further research in meaning, including pragmatic and sociolinguistic meaning, should be carried out first.

16.3. PROSODIC TYPOLOGY

In this section, a model of prosodic typology is proposed based on the prosodies of various languages which have been analysed in the framework of the AM model of intonational phonology. Comparisons across different ToBI models and AM models of intonation show similarities and differences in prosody across languages. As mentioned earlier, the types of tones and the tonal inventories proposed in these models distinguish languages in terms of their prosodic structure and, to some degree, the properties of their lexical prosody (whether a language has stress, lexical pitch accent, tone, or none of these). However, the AM model does not directly specify information about lexical prosody or timing units. For example, it does not distinguish languages which have lexical pitch accent only (e.g. Japanese) from languages which have both lexical pitch-accent and stress (e.g. Serbo-Croatian). It also does not distinguish languages which have similar prosodic structures and tonal types but differ in their timing (e.g. stress-timed English vs. syllable-timed Italian). This suggests that prosodic typology would not be complete unless we include prosodic events below the word as well as those above the word (see Hirst and Di Cristo 1998 for a similar view). When combining lexical and postlexical prosodic events, there seem to be two main aspects of variation which define prosodic typology. These are the prominence and the rhythmic/prosodic pattern of an utterance.

Prominence, as a category in prosodic typology, concerns both how the prominence of a lexical item in the language is realized prosodically (lexical prosody) and how the prominence relation among the words is realized postlexically. Although there are no agreed upon criteria for categories of lexical prosody (see Beckman 1986; van der Hulst 1999; Fox 2000 for various proposals on types of lexical prosody and their definitions), three categories of lexical prosody—tone, stress-accent, and lexical pitch-accent—are chosen in this chapter based on the phonetic realization of prosodic features, their function, and the relation to intonation (following Beckman 1986; Ladd 1996; Cruttenden 1997). A language is categorized to have a 'tone' feature if the language has 'prescribed pitches for syllables or sequences of pitches for morphemes or words' (Cruttenden 1997: 8–9), i.e., pitches having paradigmatic contrast. A language is categorized to have a 'stress-accent' feature if a certain syllable in a word is more prominent than other syllables by duration and/or amplitude, showing syntagmatic contrast. The syllable has no lexical specification of pitch but can be realized with a certain pitch pattern determined by intonation. Finally, a language is categorized to have a 'lexical pitch-accent' feature if a certain, not every, syllable of a word has lexical specification of pitch, showing syntagmatic contrast, but does not exhibit phonetic 'stress' in the sense of Beckman (1986). Therefore, unlike the stress-accented syllable, the pitch pattern of a lexical pitch-accented syllable, if realized, is fairly independent of intonation.

Data suggest that there are two ways of prominence realization at a postlexical level: culminatively by marking the head of a prosodic unit and demarcatively by marking the edge of a prosodic unit (Hyman 1978; Beckman 1986; Beckman and Edwards 1990; Ladd 1996; Venditti *et al.* 1996). Prominence is realized culminatively at a postlexical level when a syllable or word becomes prominent through a local manipulation of suprasegmental features such as pitch, duration, and/or amplitude. This is represented as the postlexical pitch accent (marked as *) in the AM model. The realization of postlexical pitch accent depends on what suprasegmental features the language employs for the realization of the lexical prosody. If the postlexical pitch accent is from the lexical pitch-accent as in Japanese, it does not change the duration or amplitude of the syllable (Beckman 1986; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988). On the other hand, if the postlexical pitch accent is from the stress-accent as in English, it does change both the duration and amplitude of the syllable. Acoustic correlates of postlexical pitch accent are language specific, and this difference will not be captured in a model of prosodic typology where languages are compared based on phonological categories.

Prominence can also be realized demarcatively at a postlexical level when the prominent word comes at a certain location in a prosodic unit (e.g. the beginning or the end), and a phrasal tone, mostly the same tone type, marks the edge of the prosodic unit. Examples of demarcative prominence are the Prosodic Word boundary tone in Serbo-Croatian (%L), the Phonological Phrase boundary tone in Bininj Gun-wok (Lp), and the Accentual Phrase boundary tone in Japanese (L%) and in Korean (Ha) (see Table 16.2 for more data). That is, the function of postlexical pitch accent in English and other West Germanic languages is preformed by prosodic phrasing in 'edge' prominence languages. For example, the function of postlexical pitch accent in English and other Germanic languages (such as marking focus or disambiguating an ambiguous string) is performed by placing words in the same or different prosodic units, i.e., prosodic phrasing, in Japanese and Korean (Ladd, 1996; Venditti *et al.* 1996; Venditti 2000). Venditti *et al.* showed that the function of pitch accent and deaccenting in English is delivered by phrasing and dephrasing in Korean and Japanese. In English, contrastive focus is realized by L+H* pitch accent followed by deaccenting, while, in Japanese and Korean, contrastive focus is realized by inserting a prosodic boundary before or after the focused word and dephrasing postfocused items (see Jun 1996, 2003*a*; Jun and Lee 1998; Ueyama and Jun 1998). Ladd (1996: 196–7) agrees with this view and notes that sentence accentuation can be seen as one manifestation of prosodic structure, and deaccenting or dephrasing are just different surface symptoms of the same deep structural effects.

Another category affecting prosodic typology is the rhythmic/prosodic pattern of an utterance. The rhythmic pattern refers to the timing unit which is smaller than the word, and the prosodic pattern refers to the prosodic unit above the word. These patterns are combined together as a group since a timing unit can be represented as a prosodic unit, and a prosodic unit has a rhythmic nature, as mentioned in Section 16.1. In stress-timed languages, rhythm is perceived through sequences of prominent and non-prominent syllables, and the prosodic unit including the prominent and non-prominent syllables is a Foot. Rhythm is also perceived when a sequence of syllables or words shows a repeating tone pattern (e.g. LLH - LLH - LLH; LHL - LHL - LHL). The grouping of syllables/words marked by a tonal pattern is a postlexical prosodic unit such as an Accentual Phrase or a Phonological Phrase. Rhythmicity will increase if the tonal pattern is the same. The rhythmic property of a postlexical prosodic unit can also be derived from the lengthening of a syllable at the end of the unit. A similar view is indicated by Cruttenden (1997: 7–21). He calls a Foot a 'rhythm group' and describes the boundary of an intonation-group as a break in the rhythm.

The rhythmic patterns derived by the postlexical prosodic unit, which I call a 'macro' rhythmic unit, would vary more than those derived by foot, syllable, or mora, which I call a 'micro' rhythmic unit. As mentioned in Section 16.1, the rhythmic constraints on the micro rhythmic patterns also seem to apply to the macro rhythmic patterns, though less strictly. That is, prosodic units at the same level tend to have a similar size and the boundary of each prosodic unit tends to occur at a regular interval. Evidence can be found in the size constraint on a prosodic unit. Studies on prosodic phonology of various languages (e.g. Nespor and Vogel 1986; Delais-Roussarie 1995; Jun 1996, 2003*b*; Selkirk 2000) have shown that a phonological phrase is constrained by the size of the unit (e.g. constraints on the minimum and maximum size of prosodic constituents).

The constraint on the size of a prosodic unit also applies to a larger prosodic unit. Even though the rhythmic nature of the unit would be weakened as the size of the postlexical prosodic unit increases, from Prosodic Word to Intonation Phrase, we can still feel a broad sense of rhythm from the distribution of Intonation Phrases. For example, when a sentence is short, it is produced in one Intonation Phrase, but when the sentence gets longer, it tends to have two Intonation Phrases with the boundary coming in the middle of the sentence (e.g. Gee and Grosjean 1983, Ferreira 1993). Similarly, in many languages, when a relative clause (RC) is short, speakers tend to produce no major prosodic break between the relative clause and the head noun, but when the RC is long and heavy, they tend to produce a major prosodic boundary between the head noun and the RC (e.g. Lovrić *et al.* 2000; Quinn *et al.* 2000; Jun and Koike 2003).

An informal observation on the size of Accentual Phrase (AP, the same level of the Phonological Phrase in Selkirk's hierarchy) and Intonation Phrase (IP) across four languages based on the short story, *The North Wind and the Sun*, shows that APs tend to have a similar size within each language and are similar in size even across languages. In Japanese, most APs have 4–5 syllables (Ueyama 1998), and in Korean and French, most APs have 3–4 syllables or 1.2 content words (Jun and Fougeron 2000). The size of an IP in each language varies more in general but becomes similar when short IPs consisting of a connective or an exclamation (e.g. *Alors*, 'Therefore', in French) are excluded. The average number of syllables within an IP ranges from 7–10 syllables in English, French, and Japanese to 12–15 syllables in Korean. The syllable count or the word count, however, could misrepresent the phenomenon because it would depend on the structure of a syllable and whether a language has function words, and if so, how it treats the function words in forming a prosodic unit. Another, perhaps a better, way to measure the rhythmic pattern of an IP would be the duration of the phrase. It is reported that the average duration of an IP is 1.5 seconds in more than one language (data from two subsamples of the Verbmobile-database and MARSEC corpus for German and English, reported by Batliner 2001; also from an interview with a French novel writer, reported in Fagyal 1995). More data should be compared to confirm the rhythmic nature of a macro prosodic unit. When doing so, we should control the type of discourse and genre and/or the speech style because these factors would affect the duration of the prosodic units.

Languages seem to differ in how an utterance is rhythmically and prosodically organized. Based on the AM model of various languages, some languages have only one prosodic unit above the word (e.g. Serbo-Croatian) while others have three (e.g. Bininj Gun-wok, Farsi). Though the rhythmic patterns and prominence realizations at the lexical level are closely related to those at the postlexical level, the postlexical prosodic patterns are not fully predictable from the lexical properties. Table 16.2 shows the prosodic features of twenty-one languages specified in two categories—the Prominence and the Rhythmic/Prosodic Unit—with each category being divided into two levels, lexical and postlexical. Lexical prominence is further divided into three categories reflecting the three features of word prosody. This arrangement enables us to represent languages which have more than one way of marking lexical prominence (e.g. both stress and lexical pitch-accent or both tone and stress). 'LPA' under the lexical prominence column stands for 'lexical pitch-accent'. Postlexical prominence is divided into two categories reflecting the two ways of prominence realization: marking the head (culminative) and marking the edge (demarcative). Head marking is achieved by postlexical or intonational pitch accent, and edge marking is achieved by the phrasal tones marking the edge of a prosodic unit (similar to or slightly higher than the Word level). When the category of prominence for a given language is not agreed upon among the researchers, '(x)' is given.

Similarly, the category 'Rhythmic/Prosodic unit' is divided into two categories, lexical and postlexical. The lexical rhythmic unit is further divided into three categories reflecting the three types of timing unit: mora-timed, syllable-timed, and stress-timed (a 'Foot' is given as the rhythmic unit of stress-timing). Though this three-way division is not as categorical as the table implies, each language is, at the moment, categorized as belonging to only one timing unit except for Italian. For languages for which there is no published data about the timing unit, a category is chosen based on consultation with native speakers and researchers working on the language

Table 16.2. Prosodic typology based on prominence and the rhythmic/prosodic unit

Prosody	**Prominence**					**Rhythmic/prosodic unit**					
	Lexical			*Postlexical*		*Lexical*			*Postlexical*		
Language	tone	stress	LPA	head	edge	mora	syll	foot	AP	ip	IP
English		x		x				x		x	x
German		x		x				x		x[a]	x
Dutch		x		x				x			x+1
Greek		x		x				x		x	x
Italian		x		x			x	(x)[b]		x	x
Spanish		x		x			x			(x)[c]	x
Portuguese		x		x			x				x
Arabic		x		x				x		x	x
Farsi		x		x	x		x		x	x	x
BGW		x		x	x			x	(x)[d]		x+1
Swedish		x	x	x				x			x
Sb-Croat.		x	x	x	x			x			x
Chickasaw		x	x	x	x			x	x[e]		x
Japanese			x	x	x	x			x	(x)[f]	x
Basque			x	x	x		x		x	x	x
French				x	x		x		x	(x)[g]	x
Bengali				x[h]	x		x		x		x
Korean					x		x		x		x
Mandarin	x	x		(x)[i]			x			x	x
Cantonese	x						x				x
Kinande	x				x		x		x		x

Notes: Spanish data is from Nibert 1999; Prieto 1999; Sosa 1999; Beckman *et al.* 2002. Portuguese is limited to European Portuguese, and the data is from Frota 2000; Frota *et al.* 2002. Arabic is limited to Lebanese Arabic, and the data is from Chahal 2001. Farsi data is from Jun *et al.* 2003. Basque is limited to the Lekeito dialect, and the data is from Elordieta 1997, 1998; Jun and Elordieta 1997; Elordieta and Hualde 2001. French data is from Jun and Fougeron 1995, 2000, 2002. Bengali data is from Hayes and Lahiri 1991; Ladd 1996; Lahiri and Fitzpatrick-Cole 1999. Kinande data is from Hyman 1990.

[a] Féry (1993) and Grabe (1998) adopt Gussenhoven's (1984) model and do not propose an ip for German. However, as noted in Chapter 3, the differences are of a theoretical rather than a typological nature.

[b] Chapter 13 (Italian) explains that not all dialects are syllable-timed. Some Southern dialects tend more towards stress-timing.

[c] Sosa (1999) and Nibert (2000) propose an ip in Spanish but Beckman *et al.* (2002) do not.

[d] The smallest prosodic unit larger than a word in BGW is a Phonological Phrase (PhP), mostly containing one morphosyntactic word. The authors (Chapter 12) claim that not all dialects of BGW show this unit.

[e] The Chickasaw AP can be larger or smaller than a prosodic word.

[f] Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988) proposed an ip for Japanese.

[g] Jun and Fougeron (2000) proposed an ip for French with a note that they need more data to confirm the unit.

[h] Lexical accent in Bengali is realized as pitch accent at the postlexical level. It does not show a stress feature as in English (Ladd 1996, 2001). However, Lahiri and Fitzpatrick-Cole (1999) found that focus clitics are lexically specified with high tone.

[i] Postlexical pitch accent in Mandarin was not proposed in Chapter 9, but a more recent proposal includes stress-driven pitch accent (Janis Fon, pers. comm. June 2003).

concerned. As can be seen in the table, the classification of the timing unit does not seem to affect the classification of other prosodic categories. Further research on the timing unit may change the category of the rhythmic unit and even the arrangement of this column.

Under the postlexical column of the Rhythmic/Prosodic unit, 'AP' is a cover term referring to the smallest rhythmic/prosodic unit above the word, whose boundary is marked suprasegmentally. This unit contains in general one morphosyntactic content word. 'ip' is also a cover term referring to a prosodic unit larger than a Word and smaller than an Intonation Phrase (e.g. the Intermediate Phrase of English, German, and Greek). This unit contains in general more than one content word and would be larger than an Accentual Phrase-like prosodic unit if a language has both (e.g. Japanese intonation proposed in Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988). '(x)' in the AP or ip column means that either the existence of the category is not fully confirmed (e.g. 'ip' in Spanish) or not found consistently across all dialects of the language (e.g. 'ip' in Bininj Gun-wok). Finally, 'IP' refers to the largest prosodic unit marked by intonation. 'X+1' under the IP column means that one more prosodic unit above the Intonation Phrase (e.g. the Utterance) has been proposed for these languages. This unit is in general not marked by intonation but by the degree of finality through phrase final lengthening and pause.

The languages are arranged in four groups following the 'traditional' characteristics of their word prosody: stress-accent languages, lexical pitch-accent languages, non-stress and non-lexical pitch-accent languages, and tone languages. In assigning the prosodic features to each language, the models described in the previous chapters are used as a reference, but if another phonological model of the same language assumes different prosodic features or structures, a superscript is given next to (x) in the table and a note is given below the table caption. For languages whose models are not described in the book, a reference is given below the table caption. It should be noted that, as in all studies on typology based on phonological categories, the categorization would be somewhat different if one adopted different models of prosody for a certain language. But, it is believed that the prosodic features of a language described in this table would not be dramatically different across different models if those models assume the same principles specified in the AM model of intonational phonology. However, the names of certain prosodic units and the number of prosodic units may differ across the models.

Table 16.2 shows that characterizing the prosodic properties of languages based on prominence and rhythmic/prosodic units allows us to observe generalizations of the relationship between the types of prominence and the types of rhythmic/prosodic unit, both at the lexical and the postlexicallevel. Generalizations found from the table are listed below.

Generalizations on Prosodic Prominence and the Rhythmic/Prosodic Unit.

(a) All languages have at least one prosodic unit above the word.
(b) In stress languages, the prominence of a word is always marked by postlexical pitch accent (i.e., marking the head of the word), but not often by marking the edge of the word.
(c) Most of the lexical pitch-accent languages mark the prominence of the word in two ways at the postlexicallevel: culminatively by marking the head of the word and demarcatively by marking the edge of the word.
(d) Languages that do not have any feature of lexical prosody mark the prominence of the word demarcatively at the postlexicallevel.
(e) Non-stress languages can have postlexical pitch accent.
(f) The number and type of rhythmic/prosodic units at the postlexical level are not predictable from the lexical rhythmic unit ofthe language, nor from the type of lexical prominence.
(g) There is no direct relationship between the type oflexical prominence and the type oflexical rhythmic unit. Also, there is no relationship between the type of postlexical prominence and the type of postlexical rhythmic/ prosodic unit. However, the edge marking ofthe postlexical prominence is predictable from the AP category in the postlexical rhythmic/prosodic unit.
(h) As 'b-d' above imply, the type of postlexical prominence is partially predictable from the type of lexical prominence.

The prosodic categorization in Table 16.2 also captures the prosodic differences and similarities across languages. As described in Beckman and Pierrehumbert (1986), English, a stress-accent language, and Japanese, a pitch-accent language, are both specified as having postlexical pitch accent and having two postlexical prosodic units. But, the table shows that, in addition to the lexical prosody and lexical rhythm, these languages differ by the way the prominence is realized postlexically: English uses head marking but Japanese uses both head and edge marking. The Accentual Phrase in Japanese tonally marks a word boundary most ofthe time (i.e., about 68 per cent ofthe rising tones atthe beginning ofan AP mark the beginning ofa word (Warner andArai 2001)), and the boundary tone is always Low (L% or wL%); but the Intermediate Phrase in English does not-the Intermediate Phrase final boundary tone can be either High or Low and is not realized on a certain syllable of a word but realized over multiple syllables between the nuclear pitch accented syllable and the last syllable of the Intermediate Phrase, i.e., post-nuclear tail (see Chapter 2).[4] The categorization also captures the

[4] An Intermediate Phrase (ip) also seems to be larger than an AP-like type when compared in terms of the number of syllables and words. When examining 20 Intermediate Phrases from the BD FM

similarities and differences between French and Korean. In both languages, the prominence of the word is not marked at the lexical level but marked at the postlexicallevel by the rising tone at the edge ofan Accentual Phrase (as mentioned earlier, the AP in these languages are about 3-5 syllables long on average and in general include one content word; Jun and Fougeron 2000; Schafer and Jun 2002). The difference between these languages lies in the fact that in French the AP final full syllable carries postlexical pitch accent while in Korean no syllable carries postlexical pitch accent. The table also shows that Japanese and Basque are similar in many prosodic features but their lexical rhythmic unit is different.

The current categorization, however, cannot capture the differences between stress languages that differ in the frequency and the type ofpostlexical pitch accent. In some stress languages like Spanish and Greek, pre-nuclear pitch accent occurs on almost all content words, and the type ofpitch accent is basically the same (e.g. L*+H for Greek) except for a few cases; while in other stress languages like English and German, pitch accent does not occur on every content word and the type varies more (see Dainora 2001 for the frequency and the type of pitch accent in English). In the former case where pitch accent occurs at a regular interval (i.e., almost every content word) with a similar type ofpitch accent, each ofthe accents would provide a cue for a word boundary, functioning similarly to the Word boundary tone in Serbo-Croatian or the Accentual Phrase boundary tone in Korean. Since the prosodic features of a language described in the AM model of intonation focus on the prosodic structure defined by distinctive pitch events and the degree of juncture, the perceptual equivalence ofword segmentation, whether it is marked by the head tone or by the edge tone of the unit, is not captured in the model. This model also does not include the category 'quantity' or length (e.g. long vowels, long consonants, heavy syllable), unlike previous discussions of prosodic typology (e.g. Hirst and Di Cristo 1998; van der Hulst 1999; Fox 2000). This is because the length distinction, if it affects the tone type, is incorporated into the tone type (e.g. Serbo-Croatian's two types of rising tone and falling tone, Japanese L vs. wL). In ToBI, the distinction in quantity is given in the Word tier as part ofthe lexical information.

In addition, the current typology does not include how languages differ or are similar in terms of the melodic category in Ladd's (2001) intonational typology (i.e., the relation between tune and meaning/function and the

News corpus, produced by three radio news speakers, Shattuck-Hufuagel found that an Intermediate Phrase included two content words on average, ranging from one to four content words (Stefanie Shattuck-Hufuagel, pers. comm. June 2003). This agrees with Ueyama's (1998) finding. Based on a short story *(The North Wind and the Sun)* read by two speakers, she found 5-6 syllables per ip (cf.3-4 syllables per AP).

realization of tunes across languages). As mentioned in Section 16.2, the relation between a tone and its meaning, or its function, is not fully studied and the currently available analyses do not provide enough data to compare across languages. Ladd's (2001) other two categories of intonational typology (accentual and prosodic) are also not captured by the current prosodic categories because Ladd's typology is concerned with the relation between prosody and morphosyntax or phonology, while the current typology is concerned with the shape (the tonal categories and the structure) of prosody itself and its realizations. As will be described in the next section, the AM model of intonation will be used as a tool to investigate the interaction between prosodic features and sub-areas of linguistics, thus helping to explore the accentual and prosodic typology in Ladd (2001).

In sum, the prosodic categorization given in Table 16.2 was chosen to capture the prosodic differences and similarities across languages described in the AM model of intonational phonology. These prosodic categories seem to be useful and significant in determining the prosodic typology of the languages described so far. In order to verify the proposed model of prosodic typology, more languages should be analysed in the AM model of intonational phonology.

16.4. ToBI TRANSCRIPTION: ITS FLEXIBILITY AND EXTENSIONS

As shown in Section 16.2, ToB! systems differ across languages, reflecting the different prosodic systems of languages. The conventions and principles assumed in the ToB! system (see Chapter 2), which are based on American English intonation, have been applied to various languages whose prosodies differ substantially. This suggests that the ToB! system is flexible without losing the integrity of the system. It is flexible to include other stress languages because the tonal inventories on the Tones tier and the break indices on the Break Index tier can change, reflecting the tonal patterns and the prosodic structure of the target language (e.g. H+L* is found only in Greek; /\H% is found only in German; the break index 3 is a juncture for an Intonation Phrase level in Bininj Gun-wok, but is a juncture for an Intermediate Phrase in German and Greek). It is also flexible to include lexical pitch-accent languages because the lexical tonal event is represented on the Tones tier in the same way as the postlexical pitch accent (e.g. H*+L). It is flexible enough to include a language which has lexical pitch-accent as well as stress-accent (e.g. Chickasaw) because it allows us to add a diacritic on a tone to distinguish lexical pitch accent from postlexical pitch accent and to label them on the same tone tier (i.e., H^{λ} for lexical pitch-accent and H* for postlexical pitch accent). This is an extension of the usage of the diacritic to mark the affiliation of the boundary tone to a different prosodic unit (e.g. H- for an Intermediate Phrase and H% for an Intonation Phrase). It is also flexible enough to include tone languages, with or without stress, because it allows labellers to create new tiers (e.g. Syllable, Stress) to tag information unique to the prosody of the language. Finally, it is also flexible enough to include a language which has no stress and no lexical pitch-accent but only has a phrasal tone and a boundary tone. This is because it allows us to have the Tones tier with no starred tone (*T) if no tone is associated with a syllable, either metrically strong or marked in the lexicon, and it allows us to expand the Tones tier to distinguish the underlying, phonological tone from the surface, phonetic tone.

One of the principles assumed in the ToBI system is that we should transcribe *distinctive* prosodic information. As mentioned in Chapters 1 and 2, the prosodic model assumed in ToBI is a phonological model, not a phonetic one. The question remains what the distinctive properties of prosody are. An apparent answer would be that a prosodic feature is distinctive if modifying that feature affects the semantic and pragmatic meaning of an utterance. But, a prosodic feature can also change the sociolinguistic meaning such as dialect variation or speech style while keeping the semantic and pragmatic meaning the same. To a research community working on the sociolinguistic meaning of prosody, the prosodic features changing the dialect identity or speech style must be 'distinctive' and should be transcribed. This suggests that there are different levels of distinctiveness, and which level of distinctiveness one should include in the transcription would depend on the interests of the research group. Similarly, distinctiveness in meaning can be realized in more than one way (cf. systemic differences in Ladd's (1996) typology). It could be done by a distinct tonal shape such as High or Rising (LH), or by pitch range manipulations such as pitch range expansion and reduction related to focus. In other words, pitch range manipulation can be phonologized to mark certain semantic or pragmatic meanings (cf. Ladd 1996; Jun and Lee 1998). Mandarin ToBI tags pitch range information on the Tones tier (e.g. %q-raise, %compress), and this is an extension of the conventional ToBI transcription because the tonal inventories on the Tones tier have always been a tone symbol (H or L and its combinations), possibly combined with some diacritics (e.g. *, -, %, !, /\, ».

Related to this is the issue of distinctive tone levels in ToBI. In the AM model of intonation, intonation contours are represented by only two tone levels, High and Low. Tonal levels other than Hand L were explained by phonological rules and phonetic realization rules such as downstep and upstep. However, in the ToBI models, downstepped High and upstepped High tones are explicitly marked by diacritics (! and /\) and they help to avoid making a transcription too abstract. This made it possible to represent four tone levels which are paradigmatically contrastive (e.g. H+ !H* vs. H+L*, H% vs. /\H% in German; H% vs. !H% in Greek).

Another extension ofthe Tones tier transcription can be found in Mandarin and Cantonese ToBI. As contour tone languages, Mandarin and Cantonese have multiple levels of pitch change within one syllable. So, instead of using Hand L, they use numerical numbers (e.g. 221, 55, 32) for the transcription of lexical tones. Cantonese tags the numerical tone on the Tones tier, but in Mandarin ToBI, the lexical tones are labelled on the Romanization tier and 'true' intonational tones are labelled on the Tones tier. The authors of Mandarin ToBI maintain that they did not include the lexical tone on the Tones tier because the realization of the underlying tone varies due to the degree of stress and tone sandhi, and the distinction between the tonal reduction and tone sandhi was not easy to resolve. For these two languages, it might be possible to add the lexical tone information on the Words tier or Romanization tier and mark the surface realizations (postlexicallevel) of the lexical tones on the Tones tier together with the 'true' intonational tones.

Another principle of the ToBI system is that the transcription conventions should be efficient. This means that we should not transcribe predictable or redundant information, either predictable from the intonation system (e.g. English nuclear pitch accent is not transcribed separately from pitch accent since it is predictable from the location of pitch accent within an Intermediate Phrase, while in Italian nuclear pitch accent is not predictable from the location of the pitch accent, and Italian ToBI labellers tag the nuclear pitch accent with a diacritic 'n' after the tone, e.g. L+H*n) or extractable from an on-line dictionary (e.g. the location of lexical stress in English) or from other tiers (e.g. the break index 4 from L% or H% on the Tones tier or vice versa).5 This suggests that we should not label the Accentual Phrase initial tones in Korean (predictable from the segment types in the Words tier) nor the Japanese Accentual Phrase initial L or wL tone (predictable from the syllable structure from the Words tier). Furthermore, we should not label Japanese pitch accent but only label the *location* of pitch accent at the postlexicallevel because there is only one type of pitch accent (i.e., H*+L)

[5] In the currently available ToBI models, German ToBI is the only one which does not label the unambiguous break indices 3 and 4 to avoid the problem of redundancy.

and this will be predictable from the intonation system ofJapanese. Similarly, if a language has a default pre-nuclear pitch accent occurring in almost all content words (e.g. Greek, Spanish), we may not need to label the pitch accent but only label non-default pitch accent.

However, not transcribing a certain value or a label due to redundancy or predictability would not necessarily increase the efficiency of the transcription. Skipping a certain label would require labellers to check other tiers or would require them to spend some time deciding if the labels were predictable or not, and this could confuse the labellers and increase transcription time. Furthermore, since all ToBI systems are 'an ongoing research program rather than a set of "rules" cast in stone' (Chapter 2), researchers often want to mark all prosodic information, whether it be redundant or not, to study the relation between prosodic features on different tiers and their phonetic realizations. Labelling information on all tiers despite the redundancy would also be useful when one is interested in examining only one type of prosodic information and not the others. For example, researchers who are interested in the degree of juncture between words relative to the morphosyntactic information of the utterance would need to look at the Break Index tier and the Words tier (or probably add a Syntax tier as described in the next paragraph), but not necessarily the Tones tier.

The ToBI systems described in this book show that a new tier can be added to describe the prosodic system specific to a language or a research community. Some tiers, like the Stress tier in Mandarin ToBI, are directly related to the realization of the tone, while others provide information about the speech (e.g. names of the dialect in the Codes tier) or information about the segment (the Phonetic Transcription tier in Chickasaw or the Prosodic Word tier in Greek). The Prosodic Word tier in Greek was added because the researchers were interested in studying the domain of segment sandhi, related to the prosodic structure. This tier will provide data for phonology-syntax mapping or prosodic phonology. All this suggests that a ToBI system is a tool with which to do research into all aspects of prosody and the interface between prosody and sub-areas of grammar. Researchers can add a Discourse tier (similar to the Japanese Finality tier) to tag a discourse structure and examine its relation to the pitch range and the degree of juncture, or add a Morphosyntax tier to study syntax-prosody relations. In addition, a tier can be split, as in the Tones tier in Korean ToBI being split into the Phonological tone tier and the Phonetic tone tier. This was done to study the distinctiveness of the phrasal tone, the realization of the underlying tone, and the tone-text alignment. Similarly, a Tone tier could also split into a Lexical Tone tier and a Postlexical Tone tier for a tone language or a language with a lexical specification of pitch for a part of the lexicon (e.g. Chickasaw, Zapotec). This would show how lexical tones are modified by or interact with an intonational tone. Finally, researchers interested in speech synthesis or recognition could add an Implementation tier, providing an interpretation of the tonal categories, which would be useful as input into the synthesizer. Adding a new tier would not decrease the integrity of the system as long as the system is used to study different aspects of prosody or the role of prosody in linguistics through the transcription of tones and break indices.

The flexibility of the system, however, could be a weakness of the system because there is no restriction on the number or types of tiers or on the types of labels or diacritics that could be added. In fact, we have seen that different ToB! systems use diacritics differently. Some use them as phonological markers and some as phonetic markers (e.g. !, <). But, this is probably unavoidable because a less developed ToB! system will include more symbols until the phonological status of the symbol is resolved by the examination of more data. The limits of the extension may also be resolved by developing more ToB! systems and using them in various types of research.

16.5. CONCLUSION

In this chapter, I have given a summary of the eleven ToB! systems described in this book. The types of tiers and break indices, types of the prosodic units, and the tonal inventories on the Tones tier reflect the prosodic structure and the intonation pattern of each language. Based on the thirteen languages described in this book and eight other languages described in the same theoretical framework, i.e., the AM model of intonational phonology, I have proposed a model of prosodic typology. The prosodic similarities and differences across the twenty-one languages are well captured by two prosodic categories, prominence and rhythmic/prosodic unit, with each category being further divided into lexical and postlexicallevels.

I have also shown how flexible the ToB! system is in incorporating many languages which vary morphosyntactically and prosodically. I have discussed how the system has been extended by some of the ToB! models and have suggested possible additional extensions. To fully understand the universal and language specific characteristics of a prosodic system, the prosodies of more languages need to be described, and to verify the model of prosodic typology proposed in this chapter, more languages should be analysed in the framework of intonational phonology.

Finally, it is suggested that a ToB! system would be an excellent tool with which we could do research on prosodic interfaces including syntax-phonology mapping, semantic/pragmatic meaning, discourse structure, sentence processing, sociolinguistics, and speech technology. It is hoped that the next edition of this book will include a revised and improved version of each ToB! system described here and will include new ToB! systems of more languages.

REFERENCES

ABERCROMBIE, D. (1967), *Elements of General Phonetics* (Edinburgh: Edinburgh University Press).

ARVANITI, A., LADD, D. R., and MENNEN, 1. (1998), 'Stability of Tonal Alignment: the Case of Greek Prenuclear Accents', *Journal of Phonetics,* 26: 3-25.

BATLINER, A. (2001), 'A Reply to "avg. length of intonational phrase"' in Prosody Discussion List (PROSODY@LIST.MSU.EDU), archive address: http://listserv.linguistlist.org/cgi-bin/wa?A2 = ind0101b&L = prosody&D = I&F = &S = &P = 88. January 2001, week2.

BECKMAN, M. E. (1982), 'Segment Duration and the "Mora" in Japanese', *Phonetica,* 39: 113-35·

—— (1986), *Stress and Non-Stress Accent* (Dordrecht: Foris).

—— (1996), 'The Parsing of Prosody', *Language and Cognitive Processes,* 11: 17-67.

——, and AYERS-ELAM, G. (1997), 'Guidelines for ToB! labelling. Version 3.0, March 1997', ms (Ohio State University).

——, DIAZ-CAMPOS, E., MCGORY, J., and MORGAN, T. (2002), 'Intonation across Spanish in the Tones and Break Indices Framework', *Probus,* 14: 9-36.

——, and EDWARDS, J. (1990), 'Lengthenings and Shortenings and the Nature of Prosodic Constituency', in J. Kingston and M. Beckman (eds.), *Papers in Laboratory Phonology* 1: *Between the Grammar and Physics of Speech* (Cambridge: Cambridge University Press), 152-78.

——, and PIERREHUMBERT, J. (1986), 'Intonational Structure in Japanese and English', *Phonology Yearbook,* 3: 255-309.

BLOCH, B. (1950), 'Studies in Colloquial Japanese IV: Phonemics', *Language, 26:* 86-125·

CHAHAL, D. (2001), 'Modeling the Intonation of Lebanese Arabic Using the Autosegmental-Metrical Framework: A Comparison with English', Ph.D. dissertation (University of Melbourne).

COOPER, W. E., and EADY, S. J. (1986), 'Metrical Phonology in Speech Production', *Journal of Memory and Language,* 25: 369-84.

CRUTTENDEN, A. (1997), *Intonation* (Cambridge: Cambridge University Press, 2nd edn.).

CUTLER, A. (1980), 'Syllable Omission Errors and Isochrony', in H. W. Dechert and M. Raupach (eds.), *Temporal Variables in Speech. Studies in Honor of Frieda Goldman-Eisler* (The Hague: Mouton), 183-90.

DAINORA, A. (2001), 'An Empirically Based Probabilistic Model of Intonation in American English', Ph.D. Dissertation (University of Chicago).

DAUER, R. (1983), 'Stress Timing and Syllable Timing Reanalyzed', *Journal of Phonetics*, 11: 51-62.

— — (1987), 'Phonetic and Phonological Components of Language Rhythm', in the *Proceedings ofXIth ICPhS* (Tallinn, Estonia, USSR), 5: 447-50.

DELAIS-RoUSSARIE, E. (1995), *Pour une approche parallèle de la structure prosodique,* Thèse de Doctorat (Univ. Toulouse le Mirail).

ELORDIETA, G. (1997), 'Accent, Tone and Intonation in Lekeitio Basque', in F. Martinez-Gil andA. Morales-Front (eds.), *Issues in the Phonology and Morphology of the Major Iberian Languages* (Washington, DC: Georgetown University Press), 3-78.

— — (1998), 'Intonation in a Pitch-accent Dialect of Basque', *International Journal of Basque Linguistics and Philology,* 32: 511-69.

— —, and HUALDE, J. 1. (2001), 'The Role of Duration as a Correlate of Accent in Lekeitio Basque', in *Scandinavia,* D. P., Lindberg, B., and Benner, H. (eds.), *Proceedings of Eurospeech 2001* (Aalberg University, Denmark, Center for Personkommunikation).

FAGYAL, *Zs.* (1995), 'Phonostylistics of Read and Spontaneous French Speech: Situation and Speaker-dependent Temporal Variation', Ph.D. dissertation (University of Paris III Sorbonne Nouvelle).

FERREIRA, F. (1993), 'The Creation of Prosody during Sentence Production', *Psychological Review,* 100: 233-53.

FÉRY, C. (1993), *German Intonational Patterns* (Tiibingen: Niemeyer).

FOUGERON, C. (1998), 'Variations articulatoires en début de constituants prosodiques de différents niveaux en français', Thèse de doctorat (Paris III, France).

Fox, A. (2000), *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals* (Oxford: Oxford University Press).

FROTA, S. (2000), *Prosody and Focus in European Portuguese* (New York: Garland).

— —, VIGARIo, M., and MARTINS, F. (2002), 'Language Discrimination and Rhythmic Classes: Evidence from Portuguese', *Speech Prosody* 2002 (Aix-en Provence: Laboratoire Parole et Langage, France).

GEE, P., and GROSJEAN, F. (1983), 'Performance Structures: A Psycholinguistic and Linguistic Appraisal', *Cognitive Psychology,* 15: 411-58.

GRABE, E. (1998), *Comparative Intonational Phonology: English and German* (MPI Series in Psycholinguistics 7) (Wageningen: Ponsen and Looijen).

— —, and Low, E. 1. (2002), 'Durational Variability in Speech and the Rhythm Class Hypothesis', in C. Gussenhoven and N. Warner (eds.), *Laboratory Phonology 7* (Berlin and New York: Mouton de Gruyter), 515-46.

GUSSENHOVEN, C. (1984), *On the Grammar and Semantics of Sentence Accents* (Dordrecht: Foris).

HAYES, B., and LAHIRI, A. (1991), 'Bengali Intonational Phonology', *Natural Language and Linguistic Theory*, 9: 47-96.

HIRSCHBERG, J., and WARD, G. (1992), 'The Influence of Pitch Range, Duration, Amplitude, and Spectral Features on the Interpretation ofL*+H L H%', *Journal of Phonetics*, 20: 241-51.

HIRST, D., and DI CRISTO, A. (1998), 'A Survey of Intonation Systems', in D. Hirst and A. Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages* (Cambridge: Cambridge University Press), 1-44.

HYMAN, 1. (1978), 'Word Demarcation', in J. Greenberg (ed.), *Universals ofHuman Language, Vol.* 2: *Phonology* (Stanford: Stanford University Press), 443-70.

— — (1990), 'Boundary Tonology and the Prosodic Hierarchy', in S. Inkelas and D. Zec (eds.), *The Phonology-Syntax Connection* (Chicago: University of Chicago Press), 109-26.

JUN, S.-A. (1996), *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure* (New York, NY: Garland).

— — (1998), 'The Accentual Phrase in the Korean Prosodic Hierarchy', *Phonology* 15/2: 189-226.

— — *(2oo3a),* 'Prosodic Phrasing and Attachment Preferences', *Journal of Psycholinguistic Research,* 32(2): 219-49.

— — *(2oo3b),* 'The Effect of Phrase Length and Speech Rate on Prosodic Phrasing', in the *Proceedings of the 15th International Congress ofPhonetic Sciences* (Barcelona, Spain), 483-6.

— —, and ELORDIETA, G. (1997), 'Intonational Structure of Lekeitio Basque', in A. Botinis, G. Kouroupetroglou, and G. Carayannis (eds.), *Intonation: Theory, Models, and Applications* (Proceedings of an ESCA Workshop, 18-20 Sept. 1997, Athens, Greece), 193-6.

— —, and FOUGERON, C. (1995), 'The Accentual Phrase and the Prosodic Structure of French', in the *Proceedings ofXIIIth International Congress of Phonetic Sciences* (Stockholm, Sweden), 2: 722-5.

— —, — — (2000), 'A Phonological Model of French Intonation', in A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology* (Dordrecht, Netherlands: Kluwer Academic), 209-42.

— —, — — (2002), 'Realizations of the Accentual Phrase in French Intonation', *Probus,* 14: 147-72, J. Hualde (ed.), a special issue on *Intonation in the Romance Languages.*

— —, and KOIKE, C. (2003), 'Prosody and Attachment Preference in Japanese: Production and Perception', in the *Proceedings of the International Workshop on Prosodic Interfaces* (Nantes, France), 27-9 March 2003.

— —, and LEE, H.-J. (1998), 'Phonetic and Phonological Markers of Contrastive Focus in Korean', in *Proceedings of the 5th International Conference on Spoken Language Processing* (Sydney, Australia), 4: 1295-8.

— —, SCARBOROUGH, R., ARBISI-KELM, **T.,** ESPOSITO, **C.,** and BARJAM, P. (2003), 'Intonational Phonology of Farsi', ms (UCLA).

KIM, J.-J., LEE, S.-H., Ko, H.-J., LEE, Y.-J., KIM, S.-H., and LEE, J.-C. (1997), 'An Analysis of Some Prosodic Aspects of Korean Utterances using K-ToB! Labelling System', *Proceedings ofConference on Sentence Processing* (Seoul, Korea), 87-92.

Korea Telecom Research and Development Group Report (1996), *A Study ofKorean Prosody and Discourse for the Development ofSpeech Synthesis/Recognition System* [In Korean].

LADD, D. R. (1992), 'An Introduction to Intonational Phonology', in G. Docherty and D. R. Ladd (eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (Cambridge: Cambridge University Press), 321-4.

— — (1996), *Intonational Phonology* (Cambridge: Cambridge University Press).

— — (2001), 'Intonation', in M. Haspelmath, E. Konig, H. E. Wiegand, and H. Steger (eds.), *Language Typology and Language Universals: An International Handbook* (Berlin: Mouton de Gruyter), Vol. 2: 1380-90.

LAHIRI, A., and FITZPATRICK-COLE, J. (1999), 'Emphatic Clitics and Focus Intonation in Bengali', in R. Kager and W. Zonneveld (eds.), *Phrasal Phonology* (Nijmegen: Nijmegen University Press), 119-44.

LEHISTE,1. (1970), *Suprasegmentals* (Cambridge, MA: MIT Press).

— — (1977), 'Isochrony Reconsidered', *Journal ofPhonetics,* 5: 253-63.

LIBERMAN, M., and PIERREHUMBERT, J. (1984), 'Intonational Invariance under Changes in Pitch Range and Length', in M. Aronoff and R. Oehrle (eds.), *Language Sound Structure* (Cambridge, MA: MIT Press), 157-233.

LOVRIC, N., BRADLEY, D., and FODOR, J. D. (2000), 'RC Attachment in Croatian With and Without Preposition', poster presented at the AMLaP Conference (Leiden).

Low, E. 1., GRABE, E., and NOLAN, F. (2000), 'Quantitative Characterizations of Speech Rhythm: Syllable-timing in Singapore English', *Language and Speech, 43/4:* 377-401.

NAKATANI, 1. H., O'CONNOR, K. D., and ASTON, C. H. (1981), 'Prosodic Aspects of American English Speech Rhythm', *Phonetica,* 38: 84-106.

NESPOR, M., and VOGEL, 1. (1986), *Prosodic Phonology* (Dordrecht: Foris).

NIBERT, H. J. (2000), 'Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation', Ph.D. dissertation (University ofIllinois, Urbana-Champaign).

PIERREHUMBERT, J. (1980), 'The Phonology and Phonetics of English Intonation', Ph.D. dissertation (Massachusetts Institute of Technology).

— —, and BECKMAN, M. E. (1988), *Japanese Tone Structure* (Cambridge, MA: MIT Press).

— —, and HIRSCHBERG, J. (1990), 'The Meaning of Intonation Contours in the Interpretation of Discourse', in P. R. Cohen, J. Morgan, and M. E. Pollack (eds.), *Intentions in Communication* (Cambridge, MA: MIT Press), 271-311.

— —, and STEELE, S. (1989), 'Categories of Tonal Alignment in English', *Phonetica,* 46: 181-96.

PIKE, K. 1. (1945), 'The Intonation of American English', in D. Bolinger (ed.), *Intonation* (Harmondsworth: Penguin), 53-83 [1972].

PRIETO, V. P. (1999), 'Review of Sosa: La entonaci6n del esponol', in *Linguistics,* 39: 1191-9·

QUINN, D., ABDELGHANY, H., and FODOR, J. D. (2000), 'More Evidence of Implicit Prosody in Reading: French and Arabic Relative Clauses', poster presented at the 13th Annual CUNY Conference on Human Sentence Processing (La Jolla, CAl, 30 March-l April.

RAMUS, F. (2002), 'Acoustic Correlates of Linguistic Rhythm: Perspectives', *Proceedings of Speech Prosody 2002* (Aix-en Provence: Laboratoire Parole et Langage, France), 115-20.

— —, NESPOR, M., and MEHLER, J. (1999), 'Correlates of Linguistic Rhythm in the Speech Signal', *Cognition,* 73: 265-92.

REMIJSEN, B. (2001), 'Word-prosodic Systems of Raja Ampat Languages', Ph.D. dissertation (Leiden University) LOT Dissertation Series vol. 49.

ROACH, P. (1982), 'On the Distinction between "Stress-timed" and "Syllable-timed" Languages', in D. Crystal (ed.), *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour ofF. R. Palmer* (London: Arnold), 73-9.

SCHAFER, A., and JUN, S.-A. (2002), 'Effects of Accentual Phrasing on Adjective Interpretation in Korean', in M. Nakayama (ed.), *East Asian Language Processing* (Stanford: CSLI), 223-55.

SELKIRK, E. (2000), 'The Interaction ofConstraints on Prosodic Phrasing', in G. Bruce and M. Horne (eds.), *Prosody, Theory and Experiment: Studies presented to Gösta Bruce (Text, Speech, and Information Technology, Vol.* 14) (Dordrecht: Kluwer Academic).

SHATTUCK-HuFNAGEL, S., and TURK, A. (1996), 'A Prosody Tutorial for Investigators of Auditory Sentence Processing', *Journal of Psycho linguistic Research,* 25/2: 193-247·

SOSA, J. M. (1999), *La entonacion del espaflOl: Su estructura fonica, variabilidad y dialectologia* (Madrid: Cátedra).

TRUBETZKOY, N. (1939), *GrundzUge der Phonologie. Traveaux du Cercle Linguistique de Prague* 7 [Repr. (1968) Gottingen: Vandenhoek and Ruprecht].

UEYAMA, M. (1998), 'Speech Rate Effects on Phrasing in English and Japanese', ms (UCLA).

— —, and JUN, S.-A. (1998), 'Focus Realization in Japanese English and Korean English Intonation', in *Japanese and Korean Linguistics,* Vol. 7, CSLI (Cambridge University Press), 629-45.

VAN DER HULST, H. (1999), 'Word Accent', in van der Hulst (ed.), *Word Prosodic Systems in the Languages ofEurope* (Berlin and New York: Mouton de Gruyter), 3-115·

VENDITTI, J. (2000), 'Discourse Structure and Attentional Salience Effects on Japanese Intonation', Ph.D. dissertation (Ohio State University).

VENDITTI, J., JUN, S.-A., and BECKMAN, M. E. (1996), 'Prosodic Cues to Syntactic and Other Linguistic Structures in Japanese, Korean, and English', in J. Morgan and K. Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition* (Mahwah, NJ: Lawrence Earlbaum Associates), 287-311.

WARNER, N., and ARAI, T. (2001), 'Accentual Phrase Rises as a Cue to Word Boundaries', a talk presented at the Annual meeting of the Linguistic Society of America (Washington, DC), 4-7 January 2001.

A Note on Recent Work on Intonational Phonology

This note for the paperback edition provides information on recent work related to Intonational Phonology or ToB!. It is followed by a list of homepages on ToB! and the intonational phonology of various languages.

Several workshops related to intonational phonology have recently been held around the world. In April 2003, a workshop on 'Boundaries in Intonational Phonology,' organized by Merle Horne and Marc van Oostendorp, was held in Lund, Sweden (http://glow.uvt.nl/Lund/phon_cal.htm). The workshop explored the correspondence of grammatical boundaries to intonational ones and the subsequent phonological/phonetic interpretation of these boundaries. The questions that were asked were: What are the relevant aspects of syntactic structure that receive intonational interpretation? Are there differences between languages in this respect, and how should these differences be accounted for? And similarly, what constitutes the inventory of elements which function to denote intonational boundaries at various levels, and how do languages differ in this respect? Presentations from the workshop have been published in January 2006 as a special issue of *Studia Linguistica,* edited by Merle Horne and Marc van Oostendorp.

Later that year, on August 3, 2003, a workshop on 'Intonation in Language Varieties-AM Approaches' was held in Barcelona as a satellite meeting of the XVth International Congress of Phonetic Sciences in Barcelona, Spain (organized by Paul Warren; http://www.vuw.ac.nzllals/staff/paul-warren/icphs/index.htm). The workshop had four themes: 1. *Determination ofphonological categories:* How can the appropriate phonological categories be established?; 2. *Systemic vs. realizational differences:* To what extent are the differences between varieties of a language, the differences in the tonal inventories for those varieties, and to what extent are they the realizational differences of the inventory? How are comparisons of varieties best served in terms of the availability of distinguishing labels; 3. *Intonational hierarchy:* How do varieties differ in their use of the intonational hierarchy (e.g., intonational phrases, intermediate phrases) and what are the implications of such differences for AM theories of intonation structure?; 4. *Stress and rhythm:* Language varieties differ in the placement of stress and rhythmic structure. What are the best means for accounting for such differences within the AM framework? Some of the papers presented at the workshop have recently been published in a special issue of *Language and Speech* (vol. 48, November 2005: Intonation in Language Varieties).

The Nordic Prosody IX Conference, held at Lund University, Sweden, in August 2004, was another meeting where works on intonational phonology were presented. Twenty-six contributions covered a wide range of aspects of Nordic prosody including not only Scandinavian languages and dialects (Danish, Icelandic, Norwegian, and Swedish), but also of related languages (Estonian, Russian, and Orkney and Shetland dialects of English). The papers were divided into the following themes: accentuation from phonetic or phonological starting-points; prosody from general linguistic perspectives, both grammar and pragmatics; prosody modeling for human-machine interaction; and other phonetic aspects of prosody, including quantity. Approximately one third of the contributions adopted the intonational phonology framework. Revised versions of the papers given at the conference were published as a book in 2006: *Nordic Prosody. Proceedings of the IXth Conference,* edited by Gösta Bruce and Merle Horne, Frankfurt am Main: Peter Lang.

About the same time (August 13-15, 2004), the 4th workshop on MAE (Mainstream American English) ToB! was held in Boston, with the theme, 'ToB! for Spontaneous Speech'. This workshop, sponsored by NSF (Grant NO.0345627), was organized by Nanette Veilleux, Alejna Brugos, and Stefanie Shattuck-Hufnagel. Twenty-one researchers who have extensive experience in ToB! labeling convened at Simmons College to discuss the implications and challenges of spontaneous speech for the ToB! transcription systems. Workshop participants identified many issues in using the ToB! system for labeling prosody in spontaneous speech and developed methods by which these issues could be investigated over time and results disseminated to the entire community. To this end, several decisions were made:

1) Revise/extend the ToB! Labeling Guide. Since the purpose of the existing Labeling Guide had been two-fold, tutorial and reference, it was decided to split these two functions into two publications: The ToB! Tutorial and the ToB! Reference Guide. The ToB! Tutorial's goal is pedagogical, i.e., it aims at introducing the system to new users, and its topics are organized in order of increasing difficulty. A draft of the tutorial is currently available online (http://anita.simmons.edu/~tobi/). The current ToB! Labeling Guide will also be revised so that it functions more effectively as a Reference Guide which is the appropriate location to maintain the historical conventions, and to more fully describe the theoretical underpinnings and implications of the ToB! system.

2) Creation and management of a website to support discussion and to post examples (http://anita.simmons.edu/~tobi/). Already-posted and proposed topics include:
   - Adding two new tiers: the Discussion and Alternatives tiers
   - Adding diacritics to label some timing/tone mismatches
   - L* accents, especially **in** sequences of other tonally marked events
   - Proposed resurrection of certain bitonal accents: H* + L, H + L*, (H+H*)
   - Discussion of metrical grids, rhythm, and rhythmic prominence for inclusion **in** the Reference Guide
   - The role ofmeaning **in** a) use ofvarious tones and tunes (e.g., the kinds of focus, presuppositions, attitudes conveyed by prosody) and b) labeling (e.g., only using L + H* for emphatic/contrastive prominence)
   - The role ofrelative size ofpitch accents (and possibly ofboundaries)
   - Criteria for boundary tones to clarify the distinction between H-L% and L-H%
   - Investigation of relatedness among Intonational Phrases
   - Investigation of 'pre-accentual lengthening'
   - Investigation offloaters: words or small phrases which might affiliate left- or rightward.
3) Creation ofa static Repository Site for ongoing work relevant to ToBI, and explore possibilities for funding a graduate student Site Manager
   - ToBI Tutorial
   - ToBI Reference Guide
   - Feedback to new users with queries
   - Ongoing discussion and comment on contributed examples, particularly of infrequent or difficult utterance transcriptions.

Finally, a workshop on 'Standard Prosody or Prosody of Linguistic Standards? Prosodic Variation **in** Grammar Writing' (http://www.let.ru.nllgep/jp/dgfs2007/main.html) is scheduled **in** Siegen, Germany, from February 28 to March 2, 2007. This workshop, organized by Jörg Peters, Margret Selting, and Marc Swerts, emphasizes the visibility of research on prosodic variation **in** the last decade, especially **in** Autosegmental Phonology and **in** spoken language research. The findings **in** these research areas challenge the view that there is a single standard prosody shared by all speakers of a standard language. Discussions will focus on how much variation is involved **in** the prosody used by speakers of national standard languages (on which grammars can be built), and, more generally, what implications prosodic variation has for grammatical description.

Homepages for ToBI and the Intonational Phonology of various languages

ToBI homepage: http://www.ling.ohio-state.edu/~tobi/
A revised ToBI Tutorial: http://anita.simmons.edu/~tobi
Cantonese: http://www.ling.ohio-state.edu/~tobi/cantonese/index.html
Dutch: http://todi.let.kun.nll
English: http://www.ling.ohio-state.edu/~tobi/ame_tobi!
German: http://www.uni-koeln.de/phil-faklphonetiklgtobi/index.html
Greek: http://ling.ucsd.edu/~arvaniti/grtobi.html
Hindi: http://www-personal.umich.edu/~jharns/hindi.html
Japanese: http://www.ling.ohio-state.edulresearch/phonetics/LToBII
Extended Japanese ToBI *(X-IT*oBI): http://www2.kokken.go.jp/~csj/public/ index.html. (45 hours ofX-JToBI-transcribed spontaneous speech is available for a small fee).
Korean: http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html
Mandarin: http://www.ling.ohio-state.edu/~tobi/mandarin/index.html
Spanish: http://www.ling.ohio-state.edu/~tobi/sp-tobi!spanish.html
Taiwanese: http://www.ling.ohio-state.edu/~tobi/taiwanese/index.html

Dissertation/Thesis/Articles on Intonational Phonology of various languages

Arabic: Chahal, D. 2001. *Modeling the Intonation ofLebanese Arabic Using the Autosegmental Framework: A comparison With English,* Dissertation, University of Melbourne, Australia.

Basque: Elordieta, G. and J. 1. Hualde. 2003. "Tonal and durational correlates of accent in contexts of downstep in Northern Bizkaian Basque". *Journal of the International Phonetic Association* 33, 195-209.

Farsi: www.ling.ed.ac.uk!teaching/postgrad/mscslp/archive/dissertationS/2002-3 / behzad_mahjani.pdf; www.stanford.edu/~rscar/index_files/Farsi%20Intonation.pdf; www.linguistics.ucla.edu/people/grads/esposito/Farsi%20Question% 20Papef2.pdf

French: Jun, S.-A. and C. Fougeron. 2002. "The Realizations of the Accentual Phrase in French Intonation", *Probus 14:147-172.*

Portuguese (European): Frota, S. 2000. *Prosody and Focus in European Portuguese,* New York: Garland.

Spanish (Porteno): www.linguistics.ucla.edu/generallMATheses/Barjam_MA 2oo4.pdf

Index