INTONATIONAL PHONOLOGY

This second edition presents a completely revised overview of research on intonational phonology since the 1970s, including new material on research developments since the mid 1990s. It contains a new section discussing the research on the alignment of pitch features that has developed since the first edition was published, a substantially rewritten section on ToBI transcription that takes account of the application of ToBI principles to other languages, and new sections on the phonetic research on accent and focus. The substantive chapters on the analysis and transcription of pitch contours, pitch range, sentence stress and prosodic structure have been reorganised and updated. In addition, there is an associated website with sound files of the example sentences discussed in the book. This well-known study will continue to appeal to researchers and graduate students who work on any aspect of intonation.

D. ROBERT LADD is Professor of Linguistics in the School of Philosophy, Psychology and Language Sciences at the University of Edinburgh. He has published extensively on various aspects of intonation in the field's leading journals, including *Language*, *Journal of the Acoustical Society of America*, and *Journal of Phonetics*.

In this series

78 HENRY SMITH: *Restrictiveness in case theory*
79 D. ROBERT LADD: *Intonational phonology*
80 ANDREA MORO: *The raising of predicates: predicative noun phrases and the theory of clause structure*
81 ROGER LASS: *Historical linguistics and language change*
82 JOHN M. ANDERSON: *A notional theory of syntactic categories*
83 BERND HEINE: *Possession: cognitive sources, forces and grammaticalization*
84 NOMI ERTESCHIK-SHIR: *The dynamics of focus structure*
85 JOHN COLEMAN: *Phonological representations: their names, forms and powers*
86 CHRISTINA Y. BETHIN: *Slavic prosody: language change and phonological theory*
87 BARBARA DANCYGIER: *Conditionals and prediction*
88 CLAIRE LEFEBVRE: *Creole genesis and the acquisition of grammar: the case of Haitian creole*
89 HEINZ GIEGERICH: *Lexical strata in English*
90 KEREN RICE: *Morpheme order and semantic scope*
91 APRIL MCMAHON: *Lexical phonology and the history of English*
92 MATTHEW Y. CHEN: *Tone Sandhi: patterns across Chinese dialects*
93 GREGORY T. STUMP: *Inflectional morphology: a theory of paradigm structure*
94 JOAN BYBEE: *Phonology and language use*
95 LAURIE BAUER: *Morphological productivity*
96 THOMAS ERNST: *The syntax of adjuncts*
97 ELIZABETH CLOSS TRAUGOTT AND RICHARD B. DASHER: *Regularity in semantic change*
98 MAYA HICKMANN: *Children's discourse: person, space and time across languages*
99 DIANE BLAKEMORE: *Relevance and linguistic meaning: the semantics and pragmatics of discourse markers*
100 IAN ROBERTS AND ANNA ROUSSOU: *Syntactic change: a minimalist approach to grammaticalization*
101 DONKA MINKOVA: *Alliteration and sound change in early English*
102 MARK C. BAKER: *Lexical categories: verbs, nouns and adjectives*
103 CARLOTA S. SMITH: *Modes of discourse: the local structure of texts*
104 ROCHELLE LIEBER: *Morphology and lexical semantics*
105 HOLGER DIESSEL: *The acquisition of complex sentences*
106 SHARON INKELAS AND CHERYL ZOLL: *Reduplication: doubling in morphology*
107 SUSAN EDWARDS: *Fluent aphasia*
108 BARBARA DANCYGIER AND EVE SWEETSER: *Mental spaces in grammar: conditional constructions*
109 HEW BAERMAN, DUNSTAN BROWN, AND GREVILLE G. CORBETT: *The syntax–morphology interface: a study of syncretism*
110 MARCUS TOMALIN: *Linguistics and the formal sciences: the origins of generative grammar*
111 SAMUEL D. EPSTEIN AND T. DANIEL SEELY: *Derivations in minimalism*
112 PAUL DE LACY: *Markedness: reduction and preservation in phonology*
113 YEHUDA N. FALK: *Subjects and their properties*
114 P. H. MATTHEWS: *Syntactic relations: a critical survey*
115 MARK C. BAKER: *The syntax of agreement and concord*
116 GILLIAN CATRIONA RAMCHAND: *Verb meaning and the lexicon: a first phase syntax*
117 PIETER MUYSKEN: *Functional categories*
118 JUAN URIAGEREKA: *Syntactic anchors: on semantic structuring*
119 D. ROBERT LADD: *Intonational phonology, second edition*

Earlier issues not listed are also available

CAMBRIDGE STUDIES IN LINGUISTICS

General editors: P. AUSTIN, J. BRESNAN, B. COMRIE, S. CRAIN, W. DRESSLER, C. J. EWEN, R. LASS, D. LIGHTFOOT, K. RICE, I. ROBERTS, S. ROMAINE, N. V. SMITH

Intonational Phonology

Second Edition

INTONATIONAL PHONOLOGY

Second Edition

D. ROBERT LADD

Professor of Linguistics, University of Edinburgh

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521861175

First published 2008
Second edition 2008

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Ladd, D. Robert, 1947–
Intonational phonology / D. Robert Ladd. – 2nd ed.
p. cm. – (Cambridge studies in linguistics ; 79)
Includes bibliographical references and index.
ISBN 978-0-521-86117-5 (hardback) – ISBN 978-0-521-67836-0 (pbk.)
1. Intonation (Phonetics) 2. Grammar, Comparative and general–Phonology.
3. Autosegmental theory (Linguistics) 4. Metrical phonology. I. Title.
II. Series.
P222.L33 2008
414'.6–dc22 2008025652

ISBN 978-0-521-86117-5 hardback
ISBN 978-0-521-67836-0 paperback

Additional resources for this publication at www.cambridge.org/9780521861175

Transferred to digital printing 2009

Dedicated to the memory of

Dwight L. Bolinger (1908–92)

Dwight R. Ladd (1921–95)

Contents

Figures

Tables

Preface to the second edition

The first edition of this book was intended partly as a monograph and partly as a readable introduction to an influential body of work on the phonology of intonation. Monographs, by their very nature, don't normally merit a second edition, but readable introductions need to be updated in order to stay useful. Prolonging the useful life of the original work in its function as readable introduction has been the principal goal of producing this revision. However, I have not resisted the temptation to tinker with the exposition in ways that I hope will make some of the monograph's original points clearer than they were the first time round. I have also rearranged the basic structure of the book somewhat, so that the distinction between pitch on the one hand and prominence and phrasing on the other emerges more clearly.

The main thing I have done to the text, in addition to correcting minor errors and updating references, is to revise and expand sections on areas of research that have been especially productive since the mid 1990s. This includes the development of the ToBI family of intonational transcription systems; research on the alignment of pitch features with the segmental string; and theoretical and experimental work on the intonational expression of focus and information structure. As in the first edition, I have included a certain amount of background material that some will find superfluous, and provided brief definitions for terms (such as 'citation form' or 'spectral tilt') that may be unfamiliar to a significant number of potential readers. In addition, I have attempted to make the basic data more accessible by providing an online appendix (www.cambridge.org/9780521678360) that contains sound files of most of the book's examples. The online appendix also includes a short guide to the acoustic analysis of intonation. So there should be enough new material here to justify republication – but this is definitely a revised version, not a brand new book.

Since the underlying premise of the first edition was that anyone who is interested in intonation should be interested in understanding the work discussed here, let me clarify something about the book's title. Intonational phonology is a *phenomenon*, an object of study: I intended the phrase 'intonational phonology'

to be parallel to, say, 'verbal morphology'. Some readers, however, seem to have assumed that I was announcing yet another new framework, and have treated the phrase as parallel to, say, 'autosegmental phonology' instead. To the extent that there is a framework here, it is the one I dubbed autosegmental–metrical, and readers who insist on locating my work – or their own – within a framework are earnestly requested not to say things like 'the framework of intonational phonology'. At the same time, though, those who are tempted to ignore the work summarised here on the grounds that it is just another formalist fad are encouraged to think again. It is true that I say rather little about the discourse function of intonation here, but that's because I regard intonational phonology and intonational pragmatics as fairly distinct topics, like phonology and pragmatics generally.

I don't expect to do a third edition, so this is the definitive version of the book. But if the vitality of the research tradition presented here continues, I do expect that by about 2020 this edition will be as out-of-date as the first edition is now. At that point somebody else can take on the task of writing a new survey.

Acknowledgements

Since this is basically a revision of the first edition, the thanks I originally recorded are still relevant, and are reprinted unchanged below. But I should also record my thanks to several people who have helped make the second edition what it is. My extended collaborations with Amalia Arvaniti, Ineke Mennen, and Astrid Schepman over the past ten or fifteen years have helped shape much of the new material in chapters 3, 4, and 5, while several years of discussions with Sasha Calhoun, Mark Steedman, and Dan Wedgwood have led to significant changes in chapters 6, 7, and 8. In addition, I am grateful to Laura Dilley, for picking many holes in the first edition, for encouraging me to produce a second edition, and especially for reading the whole revision critically when it was nearly finished. I have also benefited enormously from Peter Graff's close reading of the first edition and his many suggestions for changes and updates.

I once again thank Karen Kay for her meticulous editorial work on a near-final draft, and the many colleagues who responded to last-minute queries and requests for references and comments. I also thank Helen Barton of Cambridge University Press for not giving up on the project – despite what must be a record number of missed deadlines – and Alison Thomas for her work as copy-editor for Cambridge. Finally, I thank Catherine Dickie for preparing many of the figures, Ziggy Campbell for technical help in preparing the sound files that accompany the book, and the native speakers of the many languages discussed here who agreed to record the examples.

Acknowledgements from the first edition

It is difficult to write a book that touches on as many different topics as this one does without running up a lot of debts of gratitude, both academic and practical. This is especially true when the book's gestation period has lasted as many years as this one's has lasted. In this note I would like to make public my thanks to those without whom *Intonational phonology* would not have finally seen the light of day. None of the people named here should be blamed for any of the book's shortcomings.

I am particularly indebted to six colleagues for many detailed discussions over many years, which have been central to the development of the ideas presented here. In something like chronological order, these six are Carlos Gussenhoven, Kim Silverman, Mary Beckman, Steve Isard, Haruo Kubozono, and Alex Monaghan. I hope I have not given any of them any reason to think that they could have made better use of the time they have spent talking with me.

Several people have read one or more chapters of the book in draft form. All of them have made useful suggestions for improvement, and some of them have spent hours (or pages) in discussion with me. Alphabetically, they are Sónia Frota, Esther Grabe, Martine Grice, Carlos Gussenhoven, Dik Hermes, Caroline Heycock, Will Leben, Marina Nespor, Francis Nolan, Willem Rump, Paul Taylor, Alice Turk, and Ann Wennerstrom; I thank them all. The list of others with whom I have discussed ideas germane to this book is potentially long enough to exceed my word limit, and I apologise for not including everyone. However, I must certainly acknowledge (again alphabetically) the contributions of Amalia Arvaniti, Steven Bird, Dwight Bolinger, Bruce Connell, Anne Cutler, Heinz Giegerich, Nina Grønnum, Geoff Lindsey, Rachel Morton, Janet Pierrehumbert, Jacques Terken, Hans 't Hart, Susanne Uhmann, Enric Vallduví, Jo Verhoeven, and Ron Zacharski. Thanks are also due to the participants in my courses at the European Summer School in Logic, Language, and Information (Lisbon, August 1993) and the Summer School on Field Research and

Language Description (Cologne, September 1993), who were subjected to earlier versions of many of the ideas presented here.

I also thank many speakers of the many languages discussed in the following pages for providing me with the grammatical intuitions and (in some cases) the live examples that form the basis of some of my empirical claims. Once again, it is scarcely possible to name everyone who has helped me in this way, but I would like to thank (this time alphabetically by language name) at least the following people: Aditi Lahiri (Bengali); Enric Vallduví (Catalan); Carlos Gussenhoven and Ineke Mennen (Dutch); Delphine Dahan (French); Esther Grabe, Christa Grewe, and Susanne Uhmann (German); Amalia Arvaniti (Greek); Anna Babarczy, Mária Józsa, and Szilvia Papp (Hungarian); Martine Grice and Antonella Sorace (Italian); Sónia Frota and Marina Vigário (Portuguese); Laurenţia Dascălu-Jinga (Romanian); Natasha Kostromskaya, Wayles Browne, and Jim Miller (Russian); Wayles Browne (Serbo-Croatian); Vedia Ceranoğlu (Turkish); and Lawrence Adéwọlé (Yoruba). I explicitly take responsibility for any misuse I may have made of the information with which they provided me.

On the practical side, I first of all thank Dominic Watt, Glyn Bottrell, Andrew Ladd, Ineke Mennen, Heather King, Diane Nelson, and Karen Kay, who together put in many hours of tedious editorial, computer, and library work on my behalf. I am especially grateful to Diane for her careful last-minute work on the bibliography and indexes, and to Karen for her unerringly ruthless way with mixed metaphors and excessively clever turns of phrase. Thanks are also due to the many colleagues near and far who answered last-minute requests for facts, examples, figures, and obscure references as the book neared completion. Finally, thanks are due to Judith Ayling of Cambridge University Press (for not losing patience with my repeated assurances that the book was indeed nearing completion), and to Jenny Potts, who copy-edited the entire typescript for the Press.

Sarah Hawkins and Tom Baer rented me a spare room in their house near Cambridge from January to March 1989, where I wrote much of an early draft of the book. René Collier and Jacques Terken helped arrange for me to spend six months (March–August 1994) as a Van Houten visiting research fellow at the Institute for Perception Research in Eindhoven, and Anne Cutler invited me to spend a month (July 1995) as a visiting researcher at the Max Planck Institute for Psycholinguistics in Nijmegen. The intellectual stimulation and the uninterrupted time for thinking and writing that I found at both institutes was invaluable for the book's development.

Closer to home, I thank my family – Antonella, Andrew, Marco, and latterly Carlo – for putting up with the long hours I spent at the office during the summers of 1993, 1994, and 1995. Nor should I neglect to mention the practical assistance during those summers of my parents-in-law, Carmelo and Luisa Sorace. I should particularly take appreciative note of Luisa's cooking, without which I would probably have stayed late at the office many more times than I did. I look forward to resuming something more closely resembling normal family life now that the book is finished.

I Preliminaries

1 *Introduction to intonational phonology*

Research on intonation has long been characterised by a number of unresolved basic issues and fundamental differences of approach. For many years, these precluded the emergence of any widely accepted framework for the description of intonational phenomena, or even any general agreement on what the interesting phenomena are. Since the mid 1970s, however, several lines of research have converged on a set of broadly shared assumptions and methods, and studies on a variety of languages are now yielding new discoveries expressed in comparable terms. This emerging viewpoint – which it is perhaps only slightly premature to characterise as the standard theory of intonational structure – is the subject of this book.

As the book's title suggests, the heart of this theory is the idea that intonation *has a phonological organisation*. This idea requires some justification, since intonation sits uneasily with many ordinary linguistic assumptions. For one thing, it is closely linked to a paralinguistic vocal code: sometimes against our will, pitch and voice quality help signal information about our sex, our age, and our emotional state, as part of a parallel communicative channel that can be interpreted by listeners (even some non-human ones) who do not understand the linguistic message. Yet we know that in languages like Chinese or Thai or Yoruba it is also fairly simple to identify a small inventory of phonological elements – tones – that are phonetically based on pitch or voice quality but are otherwise quite analogous to segmental phonemes, and that these tones function alongside the more universal paralinguistic effects of pitch and voice quality. The question is thus not whether pitch and voice quality *can* have phonological structure – we know that they can – but whether intonation does have phonological structure in languages like English or French. By claiming that it does, we focus in some sense on the distinction between linguistic and paralinguistic functions of pitch, and claim that intonation belongs to the former.

Another apparent obstacle to talking about intonational phonology is that the phonetic substance of intonation somehow seems less concrete than the properties involved in consonants and vowels. In the case of pitch, instead of

the complex constellations of articulatory settings or acoustic parameters that identify a [t] or an [o], we find only a simple scale of up and down, which can differ conspicuously from speaker to speaker and occasion to occasion; somehow the phonological properties of pitch have to be defined relative to the speaker and the occasion. In the case of stress, its very definition seems to depend on a comparison between one word and another or one syllable and another: the identification of a given syllable as stressed often seems to depend on its perceived prominence relative to some other syllable that is not stressed. Understanding the ways in which these features need to be defined in relative terms will require us to go beyond the traditional paradigmatic basis of phonology – contrast based on the choice of one element rather than another from a set of possibilities – to consider syntagmatic contrasts that depend on the structural relation between one element and another in the same utterance.

These two dichotomies – paralinguistic versus linguistic, and syntagmatic versus paradigmatic – lie at the heart of most of the issues discussed in the book. Once we understand them better, I believe that it will seem as natural to talk about the phonology of intonation as to talk about the phonology of ordinary words. After that, we should eventually reach the point of being able to describe in explicit and testable terms how intonation affects the meaning and function of utterances. What follows can be thought of as a kind of report on our progress toward that goal.

1.1 Intonation

1.1.1 Three defining characteristics

We begin with a definition. Intonation, as I will use the term, refers to the use of *suprasegmental* phonetic features to convey 'postlexical' or *sentence-level* pragmatic meanings in a *linguistically structured* way. The three key points in this definition are the three italicised terms:

(1) *Suprasegmental*: I follow phonetic tradition in restricting my attention to suprasegmental features – features of fundamental frequency (F_0), intensity, and duration, according to a common definition. Although this restriction is traditional, it is not without problems, which I will only mention here. First, there is a problem of definition. Lehiste (1970) defines suprasegmentals as features of 'pitch, stress, and quantity'. The difference between her definition and the one I have given raises the more general question of the relations among physical, psychophysical, and phonetic properties. 'Stress' is clearly a phonetic property (i.e. a complex perceptual amalgam only indirectly relatable to psychophysical and physical dimensions); 'loudness' is psychophysical; 'intensity' is physical. Similar distinctions can be drawn in the case of 'pitch' and 'F_0', or 'quantity' and 'duration'. In all these cases, it is often unclear which terms of reference are most appropriately used in talking about suprasegmental phenomena. For the most part I have avoided this issue in what I have written here; in particular, I have made no attempt to distinguish rigorously between pitch and F_0. Strictly speaking, F_0 is a physical property and pitch is its psychophysical correlate, but in many contexts outside psychophysics little ambiguity arises if the terms are used interchangeably, and this accords with much recent phonetic work.[1]

The other problem with restricting our attention to suprasegmental features is that there are other phenomena that might otherwise be covered by the definition of intonation proposed here. For example, it has long been observed that many languages use segmental morphemes to convey the kinds of meanings that in other languages can often be signalled intonationally. Two obvious examples are question particles and focus particles (see König 1991); there are several reports of detailed similarities between typical intonational functions and the function of particles in certain languages, such as German (Schubiger 1965, 1980) or Russian (Arndt 1960). It may be that the functional similarity between such particles and intonation as defined here should outweigh the clear phonetic and syntactic differences. Similarly, research on sign language phonology has suggested a three-way distinction closely comparable to the lexical–intonational–paralinguistic distinction that I will attempt to justify below (see Liddell 1977; Wilbur 1994a, 1994b). If such a comparison is valid, it is clearly important not to define intonation solely in terms of phonetic suprasegmentals. As in other areas of phonology, sign language research may be able to yield important insights into what is essential about intonation in spoken language, and what is accidental. However, throughout the book I have deferred to phonetic tradition, and have excluded both particles and sign language from any detailed consideration.

1 For excellent tutorial reviews of the literature on pitch and loudness as they are relevant to the study of intonation, see Beckman 1986: chs. 4 and 5. Two other useful literature reviews on the processing and use of suprasegmental features are Shattuck-Hufnagel and Turk 1996 and Cutler, Dahan, and Donselaar 1997.

(2) *Sentence-level or postlexical*: intonation conveys meanings that apply to phrases or utterances as a whole, such as sentence type or speech act, or focus and information structure. By this definition, intonation excludes features of stress, accent, and tone that are determined in the lexicon, which serve to distinguish one word from another. For example, English *permit* (noun) and *permit* (verb) are composed of identical strings of phonemes, and distinguished by whether stress falls on the first syllable or the second; Standard Chinese *huā* 'flower' and *huà* 'speech, language' are segmentally identical, but are distinguished by the fact that the former has high level pitch ('Tone 1') while the latter has sharply falling pitch ('Tone 4'). Intonational features are, by definition, never involved in signalling such distinctions. Phonetically, of course, lexical features of stress, accent, and tone interact with intonational features in many ways. In general, however, the two types can be kept distinct in a description.

(3) *Linguistically structured*: intonational features are organised in terms of categorically distinct entities (e.g. low tone or boundary rise) and relations (e.g. stronger than/weaker than). They exclude 'paralinguistic' features, in which continuously variable physical parameters (e.g. tempo and loudness) directly signal continuously variable states of the speaker (e.g. degree of involvement or arousal). Like lexical features, paralinguistic features interact with intonational features. Unlike lexical features, paralinguistic aspects of utterances are often exceedingly difficult to distinguish from properly intonational ones, and it is a matter of considerable controversy which aspects are which, or whether such a distinction is even possible. I will return to discuss this at length at the end of the chapter, and at various places throughout the book.

1.1.2 *Pitch and relative prominence*

Formally and functionally, the phenomena covered by the three-part definition just given have two orthogonal and independently variable aspects, which we might refer to as 'pitch' and 'relative prominence'. These two aspects are illustrated by the four intonational possibilities of the simple utterance *five pounds* informally sketched in figure 1.1.

Pitch. The two pitch patterns shown are by no means the only possibilities in English, but they are clearly distinct. The 'falling' pitch pattern is the one that would normally be used in a straightforward reply to a question, for example in answer to a question like *How much does it cost?* The 'rising' pattern would normally be used to convey doubt, uncertainty, or some other 'questioning' modality: it could be used to ask for confirmation that the speaker has heard

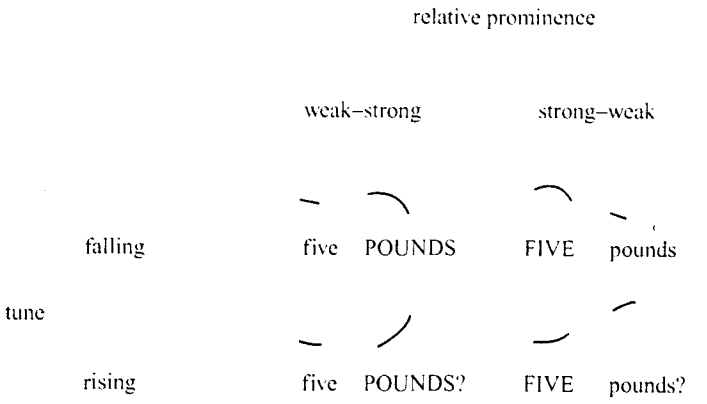


Figure 1.1. *Tune and relative prominence as two independently variable aspects of intonation, illustrated on the English phrase* five pounds.

correctly (*Did you say*) *five pounds?* Alternatively, in the relatively recent usage sometimes known as 'uptalk' or 'High Rising Terminal (HRT)', the 'rising' pattern can also be used on *five pounds* in answer to a question like *How much does it cost?* In this usage, the rising pattern would signal that the speaker is not sure of the answer, or that the price seems unreasonable, or more generally would invite feedback from the questioner about whether the price is acceptable. This shows that it is not possible to identify pitch patterns with sentence types in any simple way. It does not, however, undermine the point being made here, since the two patterns are still clearly distinct: even in the popular press notice has been taken of 'uptalk', suggesting that the distinction between falling and rising is obvious to the casual native-speaker observer.[2]

Relative prominence. The two prominence patterns are also clearly distinct. The first, weak–strong, is the 'neutral' stress pattern, used when there is no particular reason to emphasise either *five* or *pounds*, or (to put it somewhat differently) when the focus is on the phrase as a whole. This is the stress pattern that would normally be found if the phrase were used to answer a wide range of questions, like *How much does it cost?* or *What did you give him?* or *What have you got there?* The second pattern of prominence, strong–weak, focuses on *five* for contextual reasons, and would normally only be used in a discourse context where a specific number of pounds was under discussion: that

2 This use of final rises in statements is widespread in North America and in Australia and New Zealand, and appears to be spreading in Britain. An early comment in the popular press was Gorman 1993. I will return to this topic in section 3.2.3.3.

is, as an answer to a question like *Did you say four pounds?* In either pattern, it is possible to bring about perceptible gradual modifications of the phonetic 'prominence' of the individual words *five* or *pounds* by gradual changes in various acoustic parameters, adding a slight nuance of emphasis to one word or the other, but the prominence *pattern* of the utterance as a whole must fall into one of the two categories shown: either we have narrow focus on *five*, or we do not. In describing the situation this way, I am not ignoring the fact that the weak–strong pattern can be used to focus on *pounds*, for example in reply to a question like *Did you say five euros?* Rather, I am claiming that there is a clear asymmetry in the linguistic effects of the two prominence patterns, a point that will be discussed at some length in chapters 6 and 7.

The distinctions of pitch and relative prominence shown in figure 1.1 fit all three points of the definition of intonation presented above. First, of course, the features under discussion are obviously suprasegmental. Second, the meanings conveyed are clearly not lexical: the meanings of *five* and *pounds* are unaffected by the intonational changes, and the differences of pitch and of relative prominence affect the meaning of the utterance as a whole. Finally, the distinctions are linguistically structured, in the sense that we are dealing with *categories* such as rising versus falling, or weak–strong versus strong–weak. Detailed phonetic differences of prominence on the individual words occur and are meaningful, but they work within the phonological framework of the two possibilities strong–weak and weak–strong. The extent to which a categorical structure is involved in intonation is, as I said, a point of some controversy, but in these specific examples it seems fairly clear that we are dealing with sharp rather than gradual distinctions. By our definition, then, pitch and relative prominence are at the heart of intonation, and the organisation of the book's chapters is based on the centrality of these two clusters of phenomena.

Nevertheless, two points require further comment. First, in much earlier work it is often assumed that there are three main aspects to intonation rather than two. In a three-way division of intonational function, the third major function of intonation is said to be the division of the stream of speech into intonationally marked chunks ('intonational phrases', 'tone groups', and related terms). In the American phonemic tradition, for example, the three aspects were called 'pitch', 'stress', and 'juncture' (e.g. Trager and Smith 1951); juncture phonemes were supposed to be phonetically definable boundary markers of one sort or another. Halliday (1967a) states explicitly that the intonation of an utterance involves features of 'tone' (my 'pitch pattern'), 'tonicity' (part of what I am calling 'relative prominence'), and 'tonality' (the division of the utterance into tone groups). Other writers have made similar distinctions.

I do not deny of course that there are phonetic cues to the division of the stream of speech into smaller chunks, but I regard this fact as following from the existence of *phonological structure*, of the sort that has been extensively discussed in the literature on prosodic structure since the late 1970s (e.g. Selkirk 1980, 1984; Nespor and Vogel 1986; Truckenbrodt 1999). That is, I assume that utterances have a phonological constituent structure (or prosodic structure), and that the prosodic constituents have various phonetic properties, both segmental and suprasegmental. Intonation has no privileged status in signalling prosodic structure – indeed, much of the work on 'prosodic phonology' (e.g. Nespor and Vogel 1986) deals with segmental sandhi rules (rules describing phonetic adjustments at word and morpheme boundaries, such as the palatalisation that yields *gotcha* from *got* + *you*). Moreover, I assume that constituent boundaries in prosodic structure are in the first instance abstractions, not actual phonetic events: intonational features of pitch and relative prominence are distributed in utterances in ways *allowed by* the prosodic structure. In some cases this means that conspicuous phonetic breaks occur at major constituent boundaries, but this is neither the essence of the boundary nor the only factor governing the distribution of the intonational features. I will return to the issue of prosodic structure and its relation to intonational features in chapters 7 and 8.

The second point on which comment is required has to do with the relation between phonological and phonetic description. In distinguishing pitch from relative prominence and treating the two aspects of intonation as 'independent' and 'orthogonal', we are making a phonological abstraction. As can be seen from figure 1.1, there is a great deal of phonetic interaction between the two sides of the intonational coin: in short utterances like *five pounds*, the relative prominence is actually cued perceptually primarily by the pitch contour (see section 2.2). However, the fact that the pitch pattern and the prominence pattern can vary independently shows that we are dealing with two distinct phenomena: that is, in a general account of intonation, it is useful to posit an abstract prominence pattern, distinct from the pitch contours that may serve to realise it phonetically. To put this distinction in fairly traditional terms, the 'sentence stress' or 'nuclear stress' on *five* or *pounds* can be referred to independently of the 'pitch accent' or 'nuclear tone' by which it is phonetically manifested.

1.1.3 *Intonational phonology*

This brings us to the term 'intonational phonology'. Until the late 1970s there was not really any such notion, and even now it is not obvious to some intonation researchers that intonation has a phonology worth discussing. Since the publication of the first edition of this book, the actual collocation *intonational*

phonology has become more common, but as I pointed out in the preface to the present edition, some people have apparently understood it as designating a school of thought rather than a set of phenomena. It is therefore necessary to demystify this term quite explicitly.

At a minimum, a complete phonological description includes (a) a level of description in which the sounds of an utterance are characterised in terms of a relatively small number of *categorically distinct entities* – phonemes, features, or the like – and (b) a mapping between such a description and a physical description of the utterance in terms of *continuously varying parameters*, such as an acoustic waveform or tracks of the movement of the articulators. I emphasise that this characterisation of phonology is not intended to be controversial, although admittedly it has a laboratory bias that not all readers may share. I also emphasise that it is intended to apply to phonological phenomena of any sort, not just intonation. It obviously deals mostly with issues of 'postlexical' phonology and phonetic realisation, and consequently leaves out all sorts of aspects of morphophonemics or 'lexical' phonology that would be needed for a characterisation of phonology as a whole. Nevertheless, the parts it leaves out are, by the definition given in section 1.1.1, irrelevant to intonation, and it will therefore serve as an adequate notion of phonology for our purposes here.

Minimal though such a phonology may be, it is not something that is encountered very often in past work on intonation. Until the late 1970s there were two essentially separate approaches to studying intonation, which in their own way both failed to include a description that we might call phonological according to the characterisation just given. The two approaches also largely ignored each other. For want of better terms I will refer to these as the 'instrumental' approach and the 'impressionistic' approach, though – anticipating my conclusions a bit – I might also designate the two views as 'phonetic' and 'proto-phonological'. It will be useful to sketch these two approaches briefly.

The 'instrumental' or 'phonetic' tradition was that of experimental psychologists and phoneticians interested in speech perception and in identifying the acoustic cues to intonational phenomena. An excellent review of this work up to the late 1960s is Lehiste 1970. Much of this work has focused on discovering the acoustic cues to several specific intonational phenomena, in particular: (a) syntactic/pragmatic notions like 'finality', 'continuation', and 'interrogation' (e.g. Hadding-Koch and Studdert-Kennedy 1964; Delattre 1963; Lieberman 1967); (b) emotional states such as anger, surprise, and boredom (e.g. Lieberman and Michaels 1962; Williams and Stevens 1972); and (c) word and sentence stress (e.g. Fry 1958; Lieberman 1960). In none of these cases can clear understanding have been said to result, though there are some fairly general findings that are

well established, such as the fact that active emotions like anger or surprise are generally signalled by higher overall pitch (Uldall 1964; Williams and Stevens 1972), or that the duration of pauses at intonational breaks correlates well with the syntactic 'strength' of the boundary (Cooper and Paccia-Cooper 1980). But more conclusive findings seemed elusive, and fundamental uncertainty remained about such questions as the acoustic nature of stress.

The 'impressionistic' or 'proto-phonological' approach was that of linguists and language teachers who were interested in describing intonation either for practical ends (improving the pronunciation of foreign speakers of a language, in particular English) or as part of the general development of phonemic theory. This approach is represented by the work of the American structuralist school (Pike 1945; Wells 1945; Trager and Smith 1951; and others) and those of the British school (Palmer 1922; Kingdon 1958; O'Connor and Arnold 1973). Descriptions in this tradition treat intonation in terms of a small number of categorically distinct elements – pitch phonemes, nuclear tones, etc. – and in this sense may be said to be investigating 'intonational phonology'. However, in most cases the authors of these descriptions had no ambitions to go beyond data that could be gathered by traditional auditory methods and written down as impressionistic pitch curves. Moreover, for reasons that I will discuss further below, within the impressionistic tradition there were always significant disagreements about the inventory of categorically distinct elements, and there was no obvious standard of evidence for settling such disagreements.

Because of the general lack of agreement and the notable absence of instrumental evidence for impressionistic descriptions, adherents of the instrumental approach often felt that their work was somehow more rigorous or more scientific, or at the very least more complete. Such attitudes were frequently expressed by instrumental researchers during the 1960s and 1970s. For example, Lieberman and Michaels (1962: 248) state that 'most current systems of linguistic analysis of intonation seem incomplete in that they merely note gross changes of fundamental frequency, minimize the role of amplitude and phonetic variations, and entirely ignore the fine structure of the fundamental frequency ... these additional dimensions are responsible for a large fraction of the total emotional information transmitted in human speech.' More explicitly focused on methods rather than assumptions, Maeda (1976: 18) argues that a 'serious disadvantage of the studies based on auditory impressions of speech sounds is the lack of experimental means for checking whether or not the analysis is correct'. And Ohala (1975: 737f.), discussing different approaches to the study of suprasegmentals, is openly dismissive of the 'facile inventions

by taxonomic linguists' in comparison with 'what scientists of language have proven and demonstrated empirically about the behavior of speech sounds'.

As I have argued elsewhere (Ladd 1980a: ch. 6), such criticisms largely miss the point. The difference between the two approaches is not primarily one of methodology, nor one of completeness, but of theoretical assumptions. It is perfectly true that impressionistic descriptions until the 1970s made little attempt to relate their findings to instrumental work, and it is quite fair to treat this as a failing. Nevertheless, it is important to recognise that the impressionistic descriptions involve phonological categories that could *in principle* be related to instrumentally validated acoustic or articulatory parameters. More importantly, critics of impressionistic descriptions often fail to recognise that the instrumental approach also involves theoretical assumptions which can be examined and evaluated, and which do not always stand up to close inspection.

In any case, before writing an entire book on the phonology of intonation, it seems appropriate to address the views of those who question whether intonational phonology even exists. The remainder of the chapter is an attempt to do just that. In section 1.2, I discuss a theory of intonation that is 'phonological' according to the characterisation given above, but whose scientific and methodological credibility among 'instrumental' researchers is impeccable. In section 1.3, I identify and discuss two theoretical issues on which phonological and non-phonological approaches clearly disagree, and present evidence for the phonological point of view. Finally, in section 1.4, I explore some of the reasons why it has been so difficult for impressionistic descriptions and their current intellectual descendants to arrive at a widely agreed-upon set of phonological categories in the intonation of any given language.

1.2 The IPO theory of intonational structure

This section sketches the approach to describing intonation that was developed at the Institute for Perception Research (IPO) in Eindhoven between about 1965 and 1995. Though originally motivated by the search for a model of the intonation of Dutch for use in speech synthesis, the IPO approach subsequently developed into a general theory of intonational structure ('t Hart, Collier, and Cohen 1990). It served as the basis of a pedagogical description of Dutch intonation (Collier and 't Hart 1981), and was extended to the description of intonation in other languages, including English (de Pijper 1983; Willems, Collier, and de Pijper 1988), German (Adriaens 1991), Russian (Odé 1989), French (Beaugendre 1994), and Indonesian (Odé and van Heuven 1994). Though the IPO descriptive framework has now largely fallen out of use (not least because the Institute itself was redirected from pure to applied research in the mid 1990s), the IPO tradition was responsible for some important achievements. In particular, the IPO researchers were in many ways the first to make a serious attempt to combine an abstract phonological level of description with a detailed account of the phonetic realisation of the phonological elements.

1.2.1 *Phonological structure*

In the IPO approach, contours are idealised as sequences of pitch movements and connecting line segments. The model assumes that certain pitch movements 'are interpreted as relevant by the listener' and that these movements are 'characterised by discrete commands to the vocal cords and should be recoverable as so many discrete events in the resulting pitch contours, which may present themselves at first sight as continuous variations in time' (Cohen and 't Hart 1967: 177f.). Moreover, although the idealisation was not clearly stated as such in the earliest IPO work, these pitch movements are modelled on the assumption that 'the most elementary aspect of pitch variation is the difference between a relatively high and a relatively low pitch level' and that 'speech melody is characterised by the continual alternation between relatively high and relatively low pitch levels' (Collier and 't Hart 1981: 15; my translation). This notion of a basic distinction between relatively high and relatively low is a clear foreshadowing of the two-level phonologies of Bruce and Pierrehumbert, a point to which I will return in section 2.3.

Pitch movements in the IPO model are said to be of two types. The model follows Bolinger (e.g. 1958) in claiming that prominence of a particular word in a sentence is brought about by the occurrence of a pitch movement on the lexically stressed syllable of the word. Pitch movements used in this way are called 'prominence-lending'. Dutch prominence-lending pitch movements include both a rise and a fall occurring relatively early in the stressed syllable, an extremely late rise, and a kind of half-fall. These are distinguished from certain other pitch movements that are 'non-prominence-lending'. Non-prominence-lending pitch movements include a rise and a fall that occur at phrase boundaries, plus a rise and a fall that (unlike the other pitch movements in the model's inventory) may span several syllables. The non-prominence-lending pitch movements are clearly phonologically distinctive, in the sense that they may turn one contour type into a different one – for example, one of the important functions of the boundary rise is to distinguish a question from a statement – but they do not pick out a word or syllable as prominent. Cohen and 't Hart found it curious that the boundary rise, though clearly distinctive, 'need not occur in dominant words or even in prominent syllables' (1967: 189).

but such an association of distinctive pitch movements with boundaries is now well established as a characteristic of intonational phonology (see sections 2.1 and section 3.1.4).

The assumptions just outlined are well illustrated by the 'hat pattern', the best-known construct of the IPO model. This pattern consists minimally of a 'Type 1 Rise' (low to high early in accented syllable) followed by a 'Type A Fall' (high to low early in accented syllable). The two movements may occur as part of the same accent, in which case we have a 'pointed hat', or as separate accents, in which case we have a 'flat hat'. The stretches of contour preceding the rise and following the fall – and the stretch in between the two in the 'flat hat' – are idealised as straight line segments notated Ø (for the upper line) and 0 (for the lower). All these aspects of the hat pattern can be seen in figure 1.2.

The hat pattern can be represented abstractly as

(1) (0) 1 (Ø) A (0),

that is, as an obligatory 1 and an obligatory A, with optional stretches of Ø and 0 on any non-prominent syllables. *This is the phonological description.*[3] It is expressed in terms of categorically distinct entities (Type 1 Rise, etc.) that occur, in sequence, at well-defined points in the utterance. It abstracts away from differences that arise because the contour is applied to utterances of different lengths: it makes it explicit that the two contours in figure 1.2 are specific realisations of the same abstract linguistic unit, realisations that differ in predictable ways just as allophones of a phoneme differ predictably. In theory, the phonological formula in (1) tells us all we need to know about the physical properties of the hat pattern for 'higher-level' linguistic purposes: a linguist interested in the syntactic and pragmatic uses of the hat pattern would not need to be concerned with continuous F_0 parameters, only the phonological formula.

1.2.2 *Phonetic realisation*

As I noted above, a complete phonological description does not consist of abstract formulas alone, but must also specify how the abstract formulas are realised; that is, it must describe the mapping from the categorical phonological elements to the continuous acoustic parameters. The IPO researchers devoted considerable effort to this task, and the theory was used successfully as the

3 In the standard IPO notation, the pointed hat is actually represented as *1&A* and the flat hat as *1 A*. In effect, the stretches of Ø and 0 are regarded as an option that is always available, while the absence of a stretch of Ø (i.e. the occurrence of 1 and A on the same syllable) must be specified.

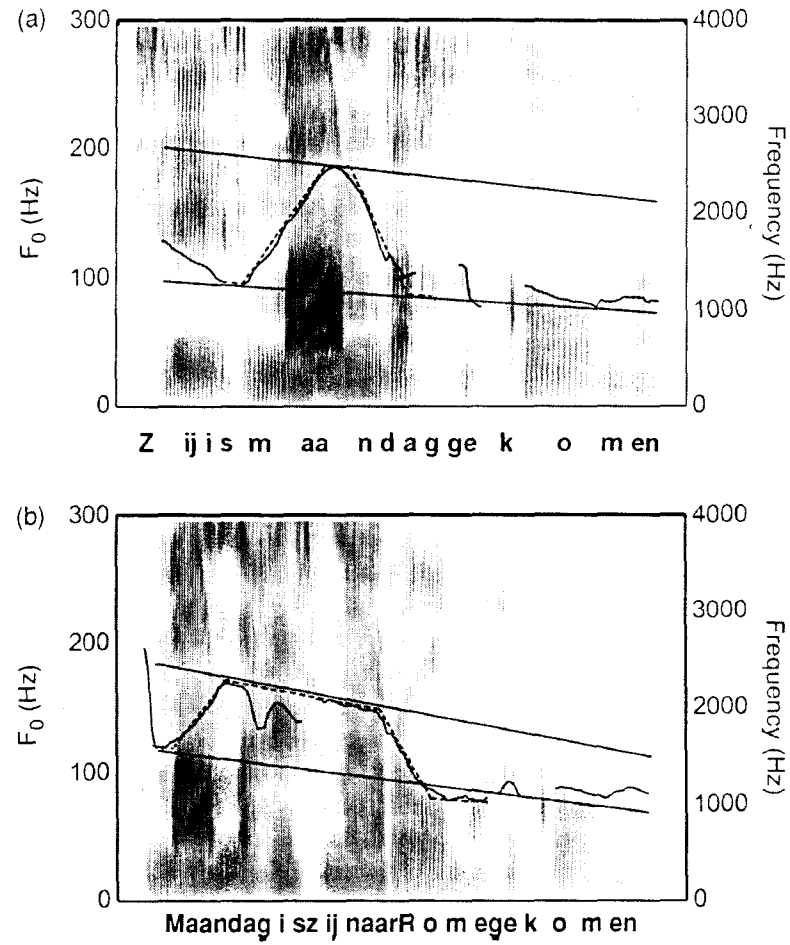


Figure 1.2. *The hat pattern in the IPO model of intonation. Panel (a) shows the 'pointed hat', with the Type 1 Rise and Type A Fall on a single accented syllable, applied to the Dutch sentence* Zij is maandag gekomen *('She came on Monday'). Panel (b) shows the 'flat hat', with the Type 1 Rise and Type A Fall on separate accented syllables, applied to the Dutch sentence* Maandag is zij naar Rome gekomen *('On Monday she came to Rome'). The superimposed straight lines show the gradually declining topline and baseline.*

basis of a model for synthesising Dutch intonation contours by rule. All the phonological elements of the model are described in considerable phonetic detail; for example, there are phonetic specifications of how pitch movements are aligned with respect to the stressed syllable. These phonetic descriptions are based on experiments that attempted to set the limits of perceptible variation – that is, how different can two pitch movements be physically and still count as perceptually the same? For more detail on this entire aspect of the IPO programme, the reader is referred to 't Hart and Cohen 1973 and 't Hart, Collier, and Cohen 1990.

Perhaps the best-known feature of the IPO approach to phonetic realisation is the notion of *declination*. This term was coined by Cohen and 't Hart (1967: 184) to describe the downward trend of F_0 observable in many utterances, and has since been adopted by a wide variety of authors. In the IPO model, declination refers specifically to the trend of the top and bottom lines that define the limits of the local pitch movements – the lines notated Ø and 0 in the abstract phonological formulas that represent contours. This means that even when nothing is 'happening' phonologically in the contour, F_0 continues to go down slightly; it also means that a pitch movement at the beginning of a phrase will be higher than the same pitch movement later in the phrase. This aspect of the model can also be seen in figure 1.2.

A considerable amount of research has gone into discovering the phonetic details of declination within the IPO framework, and the notion of declination itself has been studied extensively by investigators with a wide variety of backgrounds, interests, and theoretical assumptions. Work within the IPO theory includes 't Hart 1979, Cohen, Collier, and 't Hart 1982, and Terken 1991; reviews from different perspectives include Vaissière 1983 and Ladd 1984, 1993c. I will return to discuss this issue again in section 2.4.

1.2.3 *Key theoretical assumptions of the IPO approach*

What makes the IPO approach phonological? There are two important properties of the elements of an IPO description that distinguish them from the analytical constructs of most other 'instrumental' work on intonation. These are *phonetic definition* and *sequential structure*.

Sequential structure is obvious: an IPO description of a contour consists of a string of discrete intonational elements. To be sure, the significance of this fact is disputed. In particular, Grønnum has suggested (e.g. Thorsen 1985) that my emphasis on the sequential structure of the IPO work is misleading: she considers declination to be an equally central feature of the IPO theory, and points out that the concept of declination has been widely adopted in many

instrumental phonetic models of intonation that do not share any corresponding phonological abstractions. I do not think that this objection is valid, however. The distinction between the abstract elements and the details of their phonetic realisation is discussed very explicitly in 't Hart and Collier 1975. The abstract description clearly involves only a string of events in sequence, and declination is treated as a matter of phonetic detail, not as part of the abstract description. Declination in the IPO model certainly resembles the 'phrase components' of numerous non-phonological studies (about which more shortly), but it is part of the phonetic background, not something that is moulded to communicative ends.

As for phonetic definition, this is intended to mean that the elements of the IPO system – Type A Fall, Type 2 Rise, etc. – are defined in terms of phonetic characteristics rather than semantic ones. *Meaning or function plays no role in the analysis.* The elements are identified solely on the basis of the fact that they are perceptually distinct from other elements; and are defined solely in terms of the phonetic properties that distinguish them. There is no necessary reference to any supposed functions of intonation, and terms like 'continuation rise' are seldom used. As 't Hart and Collier (1975: 254) put it: 'we have deliberately restricted ourselves to the *melodic* aspects only. So far we have not said anything about the *functional* aspect of intonation, for example we cannot offer any explanation for the motives of the speaker that lead to the choice of a particular intonation pattern.'

Yet the question still remains: what makes the IPO work phonological? Why are sequential structure and phonetic definition crucial concepts in intonational phonology? This question has a certain amount of force, because the IPO researchers themselves tended to present their own work as being basically about speech perception, not phonology. But the significance of sequential structure and phonetic definition becomes clear when we consider segmental analogues of what the 'phonetic' approach to intonation involves.

First consider phonetic definition. In segmental phonetics, instrumental research is devoted to studying the physical cues to properties like voicing or vowel quality or nasality. Phoneticians do not try to study the physical cues to properties like plurality or verb aspect or negation – it seems obvious that it would be pointless to do so. The segmental categories investigated by instrumental phonetics are *phonological*, not lexical or grammatical. Yet one of the characteristic features of traditional instrumental research on intonation is that in many cases it attempts to identify direct physical correlates of meanings or linguistic functions, such as happiness or contrast or finality. The anomaly is seldom commented on or even recognised, but this anomaly is precisely what

the IPO model avoids when it deals in 'Type 2 Rises' rather than 'continuation rises'.

Now consider sequential structure. As an object of phonetic investigation, the 'F_0 contour' is actually quite problematic. Phoneticians who try to describe the shapes of F_0 contours would never treat 'second formant contours' in the same way. In segmental phonetics, it is taken for granted that the course of the second formant through an utterance is a function of the sequence of segmental sounds of which the utterance happens to consist. The second formant – like other formants – is often modelled phonetically as a sequence of phonologically specified targets or steady states with transitions between them. These are the assumptions that the IPO approach brings to the task of modelling F_0: the 'F_0 contour', like the 'second formant contour', arises from a sequence of categorically distinct elements.

Finally, of course, we come to the most important question: who is right? It is all very well to identify the underlying theoretical assumptions of the IPO approach, and all very well to argue that these assumptions are in line with the great body of instrumental phonetic work on segmental properties of speech; it remains the case that the IPO assumptions could be wrong. In the next section, therefore, I discuss the evidence for describing intonation in terms of strings of elements that are not defined in terms of their function.

1.3 Evidence for phonological assumptions in describing intonation

1.3.1 Intonational meaning in experimental research

As I just noted, segmental phonetic research almost never looks for things like 'acoustic correlates of negation', but in intonational work comparable quests are commonplace. 'Acoustic correlates' have been sought for a variety of meaningful aspects of utterances, including surface constituent structure, the discourse status of referring expressions, and speaker emotion and attitude. For the most part, the authors of such studies make no attempt to identify phonological categories. Instead, they simply take a set of intonational functions for granted, and assume that the most appropriate description of how these functions are expressed is in terms of the continuously varying parameters of speech – in particular, the suprasegmental parameters of F_0, duration, and intensity.

Good examples of this approach are found in the work of William Cooper and his associates. For example, in their study of cues to surface constituent structure, Cooper and Paccia-Cooper (1980) assume that 'boundary strength' (defined in terms of the depth of the node in the syntactic tree that dominates the words on either side of the boundary) is directly reflected in various acoustic

parameters, such as pause duration and the depth of F_0 valleys in the 'fall–rise' patterns;[4] related work is reported by Cooper and Sorensen (1981). In the same vein, Eady and Cooper and their collaborators (Cooper, Eady, and Mueller 1985; Eady and Cooper 1986; Eady *et al.* 1986) attempt to identify the acoustic correlates of notions like 'focus', 'contrast', and the distinction between questions and statements. Their procedures treat these notions as independent variables, and the various acoustic parameters they measure (such as duration and peak F_0 of focused word) as dependent variables. They take it for granted that functions like focus will have acoustic correlates, as the title of one of their papers ('Acoustical characteristics of sentential focus') makes clear. This general outlook still has plenty of adherents, the most notable recent work in this tradition being that of Yi Xu and his colleagues (Xu 2005; Xu and Xu 2005). Once again, a recent title ('Speech melody as articulatorily implemented communicative functions', Xu 2005) succinctly expresses the underlying theoretical assumptions.

By directly relating phonetic detail to categories of meaning, the approach taken in these studies presupposes that intonation is unlike the rest of language, because it has no place for a phonological level of description. This presupposition is implicit, of course. Intonational phonology is simply not an issue in much of this research. It seems fair to say that in virtually all of these studies the main motivation for treating suprasegmental features in terms of continuous physical parameters is laudable: to avoid the seemingly unverifiable speculation of 'impressionistic' work, and to permit the use of familiar parametric statistical approaches. But the use of rigorous methods is no guarantee of useful findings if the conceptual foundations of the research are flawed.

The absence of any attention to phonological structure was discussed at some length by Pierrehumbert and Liberman (1982: 691) in their review of Cooper and Sorensen (1981). They comment on Cooper and Sorensen's treatment of the relation between syntactic boundary strength and the depth of F_0 valleys as follows:

> The consensus of the linguistic sources cited is that the basic fall–rise patterns fall into two categories. One pattern, which has a fall to the bottom of the speaker's range, occurs at the boundary between one intonation phrase and the next. The other, which involves a less extreme fall, occurs within a single intonation phrase ... At a given syntactic boundary, an intonational boundary

4 Note that the 'fall–rises' in the work of Cooper and his associates are not usually the same as those of the British tradition.

> is typically optional. Its probability of occurring is influenced by phrase length, speech rate and style, and the information structure of the discourse ...
>
> Cooper and Sorensen present a different picture, although they do not note that it differs from that developed by linguists. Under their account, the depth of the valley in fall–rise contours varies continuously, reflecting syntactic boundary strength ... Unfortunately, the experimental data presented to support Cooper and Sorensen's position are compiled in a way that cannot discriminate between it and the alternative ... [Their] procedure would give a reasonable picture of data that varied continuously, but it would also make data that fall into two categories with varying probability look as if they varied continuously.

Pierrehumbert and Liberman's criticism, in short, is that the constructs of linguistic analyses of intonation have simply not been given a chance to demonstrate their validity.

More or less the same point is made by Arvaniti, Ladd, and Mennen (2006) in their critique of Xu's purely parametric approach to intonational meaning. They present instrumental evidence from the interaction of focus and question intonation in Greek, showing that, on its own, Xu's approach is unable to make sense of this interaction unless it also incorporates some notion of 'tonal target', a phonological abstraction that accompanies the linguistically most prominent word in the utterance. More generally, they argue that an empirically adequate description of the phonetic correlates of focus cannot be based universally and exclusively on the manipulation of phonetic realisation parameters, but in at least some languages must make reference to local phonological events that occur at specific points in the utterance.

This brings up a further point: we are not dealing only with philosophical foundations, but with fairly clear assumptions that can be made explicit and used as the basis of testable experimental questions. Whether we should adopt a 'phonological' approach to intonation is not primarily a matter of taste, but an *empirical question*. Nothing in the phonological approach requires us to use the impressionistic methods of half a century ago, just as nothing in segmental phonetics requires us to limit ourselves to IPA transcription. It is possible to work with the ideas of intonational phonology and still do methodologically rigorous instrumental and experimental work. When we do so, we often find that hypotheses based on putative phonological categories stand up to empirical tests, and lead to new insights and findings.

A second example of the assumption that phonology is irrelevant to studying intonation is provided by studies of the effects of discourse structure – such as the 'newness' or 'givenness' of a referring expression in a specific context – on the phonetic realisation of individual words. Several early studies simply

assumed that it is appropriate to look for direct acoustic correlates of the 'givenness': for example, Fowler and Housum (1987) and Shields and Balota (1991) found that 'second mentions' of nouns in a discourse were shorter than 'first mentions', while Koopmans-van Beinum and van Bergem (1989) found spectral and F_0 effects of givenness but no durational effects. To be sure, these are in the first instance psycholinguistic studies, concerned primarily with the intelligibility of referring expressions when presented out of context. The acoustic properties of the referring expressions are measured as part of an effort to explain the observed differences of intelligibility. Nevertheless, these studies fit the mould discussed in this section, in the sense that the acoustic properties in question are treated as direct correlates of givenness. In particular, none of the studies makes any attempt to control for the presence or absence of *pitch accent*. They proceed without assuming any distinction between 'accented' and 'unaccented', and therefore treat the acoustic differences that they find as a direct reflection of discourse status.

This neglect of pitch accent was called into question by Hawkins and Warren (1991), who performed the same sort of study as Fowler and Housum (1987) (i.e. testing the intelligibility of words excised from context). Hawkins and Warren showed that accentedness, not givenness per se, was the main source of differences in intelligibility. Duration correlates with accentedness, and accentedness correlates with givenness, but the continuous acoustic properties measured by Fowler and Housum are properly seen as a reflection of accentedness, and hence only indirectly of givenness. In fact, more recently, Bard *et al.* 2000 have shown that, when accent is duly controlled for, there is still a direct effect of givenness on the duration and intelligibility of referring expressions. Moreover, it appears that accentedness is not especially well correlated with givenness anyway, in the sense that words may be accented in many contexts even if they refer to given referents (e.g. Bard and Aylett 1999; Swerts, Krahmer, and Avesani 2002). That is, it appears both possible and useful to distinguish categorical or linguistic reflexes of givenness, such as accentedness, from paralinguistic or continuous correlates, such as duration and precision of articulation.

For a more extended illustration of the potential usefulness of phonological constructs in explaining the function of intonation, consider a series of experiments on intonation and emotion by Scherer and his collaborators (Scherer, Ladd, and Silverman 1984; Ladd *et al.* 1985; Ladd, Scherer, and Silverman 1986). This is an area of study in which 'instrumental' approaches have been widely adopted. In particular, there are many studies in which investigators identify a list of emotions in advance and then set about trying to identify

acoustic parameters that correlate with the chosen emotions (e.g. Lieberman and Michaels 1962; Kramer 1964; Apple and Hecht 1982; Ofuka *et al.* 2000). Instead of following this procedure, Scherer, Ladd, and Silverman (1984) based testable hypotheses on two different sets of assumptions broadly corresponding to the two approaches to intonation that we have been discussing.

Scherer, Ladd, and Silverman refer to these two approaches as the 'covariance view' and the 'configuration view'. These terms are based on statistical notions. The covariance view involves the claim that there are direct acoustic cues to emotional messages, independent of any linguistic message being expressed, and that these cues are continuous or parametric variables that *covary* with the message. The configuration view is the idea that intonation has a linguistic structure involving contrasting categories such as rise and fall, and that in statistical terms the emotion is conveyed at least in part by *configurations of category variables*. In the terms we are using here, the covariance view is the one underlying most instrumental work, while the configuration view presupposes a notion of intonational phonology.

In order to evaluate the validity of these two views, Scherer, Ladd, and Silverman asked judges to rate the emotional force of a set of ordinary utterances. The utterances were all questions, and were all taken from recordings of spontaneous speech. One set of conditions was designed to test the assumptions of the 'covariance' or non-phonological approach. In some of these conditions, the utterances were modified in various ways to render the words unintelligible, and in one condition the judges rated only transcripts of the sentences. Not surprisingly, there was good agreement among the judges on the emotional force of the modified utterances, and little agreement on the transcripts. This part of the experiment thus shows very clearly that some of the emotional message of an utterance is indeed non-phonological, or (in Scherer, Ladd, and Silverman's terms) works according to the assumptions of the covariance view.

In a subsequent part of their study, however, Scherer, Ladd, and Silverman asked judges to rate the sentences in their original form, without any acoustic manipulations, and analysed the results looking for categorical effects. Such effects appeared unambiguously. Even on a very crude phonological categorisation of the intonation contours by final pitch movement (boundary rise vs boundary fall), and a categorisation of the utterances as either yes–no or WH (question-word) questions, Scherer, Ladd, and Silverman showed that the judgements were affected by certain combinations (or, in their terms, 'configurations') of categories. For example, yes–no questions with a final fall – but not any of the other three combinations – were judged strongly 'challenging'. In the same way, there was a strong interaction of contour type and question type in the extent to which utterances were rated 'agreeable' and 'polite': rising yes–no questions and falling WH questions (i.e. the combinations of intonation and sentence type that are commonly supposed to be normal or 'unmarked') were rated high on these scales, while falling yes–no questions and rising WH questions were rated low.

Scherer, Ladd, and Silverman's findings suggest that, in order to understand how intonation conveys emotion, it is not enough to look for continuous acoustic variables that directly signal the strength of some emotional message. In addition, it is necessary to consider the categorical presence or absence of certain intonational elements at specific points in the contour (e.g. boundary rise vs boundary fall), in conjunction with other categorical linguistic properties of utterances (e.g. question type). Experiments that ignore the phonological structure of the suprasegmental cues may simply fail to address part of the problem.

1.3.2 *Sequential structure and superposition in modelling F_0*

A more specific point of contention between phonological and non-phonological approaches to intonation is the question of what I have elsewhere called *overlay* models of F_0 (Ladd 1988). Overlay or superposition models treat the linguistic pitch contour as if it were some sort of complex function, which can be decomposed into simpler component functions. Many descriptions of intonational systems of individual languages are expressed in such terms, with local 'bumps' (e.g. for accent on a prominent word) overlaid or superimposed on global shapes or slopes (e.g. for the distinction between a statement and a question). In some overlay models the components are organised hierarchically, so that there can be, as it were, bumps on shapes on slopes. Bolinger has expressed the basis of this view metaphorically (1964, reprinted 1972a:19f.):

> The surface of the ocean responds to the forces that act upon it in movements resembling the ups and downs of the human voice. If our vision could take it all in at once, we would discern several types of motion, involving a greater and greater expanse of sea and volume of water: ripples, waves, swells, and tides. It would be more accurate to say ripples *on* waves *on* swells *on* tides, because each larger movement carries the smaller ones on its back ... In speech ... the ripples are the accidental changes in pitch, the irrelevant quavers. The waves are the peaks and valleys that we call *accent*. The swells are the separations of our discourse into its larger segments. The tides are the tides of emotion.

Mathematical implementation of the metaphor was apparently first attempted by Öhman (1967), but is perhaps best known in Fujisaki's F_0 model (e.g.

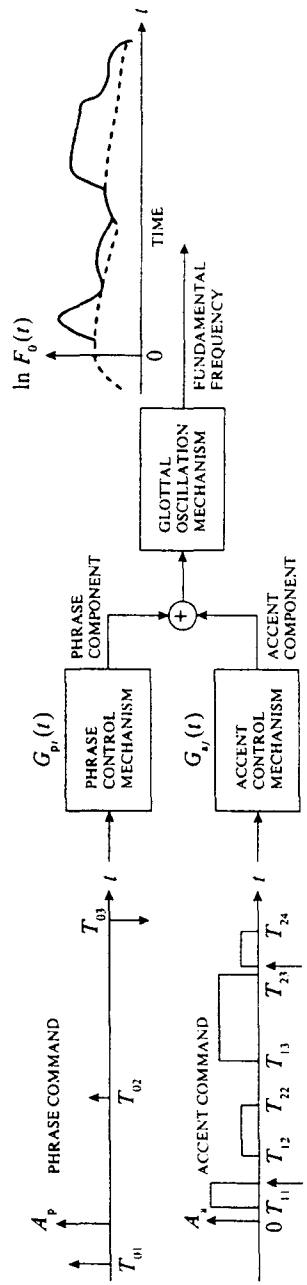


Figure 1.3. *Basic features of Fujisaki's model of intonation. The output of the accent component (a string of rise–level–fall sequences of varying height and duration, lower left side of figure) is added to the output of the phrase component (a series of one or more impulses, upper left side of figure). The impulse response decays asymptotically and the accent commands are smoothed, so that the result of adding the two components is a detailed model of a pitch contour (right side of figure).*

Fujisaki and Hirose, 1982; Fujisaki 1983). Fujisaki's model posits two components to F_0: a *phrase component* and an *accent component*. The phrase component is modelled as an impulse response: graphically, it rises rapidly to a peak and then decays exponentially towards an asymptote. The accent component is modelled as a step function, which creates a string of steps up and steps down that represent the local rises and falls of pitch at accented syllables. The step function is smoothed by the addition of a time constant, and is then added to the phrase component to create the contour. This is shown in figure 1.3.

The overlay approach, like the metaphor on which it is based, is intuitively appealing, and is almost certainly the most suitable model of certain aspects of F_0. For example, so-called microprosody – the local perturbations in the F_0 contour that are due to the effects of certain segment types – can best be seen as local 'ripples' on the contour that results from the speaker's linguistic intentions.[5] Similarly, the 'tides of emotion' that raise or lower the overall level of the voice can also perhaps best be factored out of the description of the contour. But most overlay models do not deal in detail with either microprosody or overall level. Instead, their main concern is the relation between local F_0 events (pitch movements on individual syllables or words due to accent and lexical tone) and F_0 trends that extend over somewhat larger domains (such as 'declination'). This is true of Fujisaki's model, and of quite a number of others, including the work of O'Shaughnessy and Allen (1983) on English; Grønnum (e.g. Thorsen 1980, 1985) on Danish; and Gårding and her colleagues (e.g. Gårding 1983, 1987) on various languages, including Swedish and Chinese. In all these models, the emphasis is on the superposition of contour shapes for short domains (sentence or accent group) on contour shapes for longer domains (phrase or utterance). Microprosody and overall raising of the voice are considered only in passing. I will therefore focus the following remarks on the notion of the 'phrase component' (or its equivalent) and the way it is used in overlay models generally. Quite apart from the general question of whether it is sensible to attempt to model an overall contour shape (see my comments on the 'second formant contour' in the previous section, p. 18), there are

5 The principal microprosodic phenomena are: (a) the intrinsic F_0 of vowels (see Lehiste 1970: 68–71; Whalen and Levitt 1995); (b) rapid F_0 movements in the vicinity of obstruents (see Lehiste 1970: 71–4; Reinholt Petersen 1986); and (c) gaps in the contour during periods of voicelessness. It seems likely that these phenomena behave acoustically and perceptually like local perturbations of an idealised underlying F_0 course (e.g. Silverman 1986, 1990; Kohler 1990; Whalen *et al.* 1990). There seems little doubt that an overlay model is the best way to treat them in generating F_0 for synthetic speech (see Hirst 1983; Beckman 1995). Further detail is given in the online appendix.

empirical and theoretical problems with phrase components that have never been successfully dealt with.

The most conspicuous problem with these models is that they have difficulty providing a quantitative definition of the phrase-level components that they presuppose. Even something as apparently straightforward as the 'slope of declination' has resisted quantitative characterisation. In Vaissière's words, 'declination is a general tendency easily detected from a visual inspection of relatively long stretches of F_0 curve, but calculation of the exact rate of declination is a difficult task' (1983: 56). Components with more complex shapes, which are posited in some overlay models, are correspondingly more difficult to define. For example, in the model proposed by Gårding (1983), a 'grid' is fitted to the tonally or accentually determined 'turning points' of the contour, as shown in figure 1.4. The shape of this grid is then taken as an indication of the global properties of the contour, and is related to, for instance, the distinction between questions and statements. As a rough expression of the relationship between global and local properties of contours, Gårding's model, like overlay models in general, makes intuitive sense. But if the 'grid' is to be something more than an impressionistic smoothing of graphically presented data, then it will have to be defined as a mathematical function. Moreover, if the mathematical function is to have any predictive value, it will have to generate a variety of distinct realisations (e.g. for utterances of different length); and, most importantly, it will have to provide a common characterisation of contour shapes that are perceived to have the same linguistic function. For example, if Chinese questions and statements are distinguished by the slope and width of the grid, then the function will have to provide a quantitative characterisation that distinguishes question grids from statement grids, so that the model can be tested against empirical data. As far as I know, Gårding never attempted any such definitions for any of the profusion of grid types posited in her model.

One specific aspect of the quantitative definition problem is the problem of advance planning or *lookahead*. Part of the reason it is difficult to provide quantitative descriptions of F_0 trends is that in general such trends are quite plainly dependent on the length of the domain to which they apply. Specifically with respect to declination, it has been found in numerous studies that the longer the domain over which declination can be observed, the less steep the declination slope. (An alternative way of stating this is that the amount of F_0 drop across a declination domain is relatively constant, irrespective of the length of the domain.) This finding emerges very clearly from Grønnum's work (especially Thorsen 1986), and is incorporated into models of declination in the

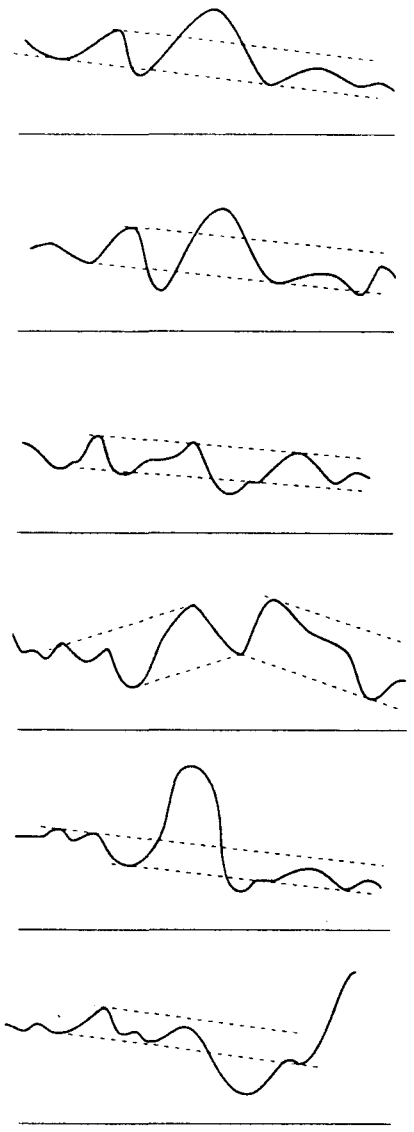


Figure 1.4. *The 'grid' in Gårding's model of intonation. One or more straight lines are fitted to most of the local minima and most of the local maxima of a pitch contour to give an approximation of its overall range and direction. The figure shows grids applied to several different renditions of the Swedish sentence* Hon gick inte och la sej *('She didn't go to bed'). From Gårding 1983. See further figure 4.3.*

IPO tradition (e.g. 't Hart 1979). This means that any speech production model – or any F_0 generation model for speech synthesis – must know in advance how long a declination domain is in order to start the declination going at the appropriate slope. There is some reason to think that this degree of lookahead may be psycholinguistically implausible; this problem has been discussed at some length by Liberman and Pierrehumbert (1984: 220ff.).

Both these problems are dealt with in Fujisaki's model. There is only one type of shape for the phrase component, and it is generated as a decaying impulse response, so there is no need for lookahead. In fact, it is reasonable to regard Fujisaki's model as a possible phonetic realisation model for a sequential phonological description: several people have suggested to me that there is no deep incompatibility between Fujisaki's approach and a sequential phonology. Both the accent commands and the phrase commands 'happen' at specifiable times, like the string of events in the IPO model. The basic compatibility can be seen in Möbius's work on German intonation, in which many of the assumptions of sequential intonational phonology are integrated with a Fujisaki-style realisation model (Möbius 1993; Möbius, Pätzold, and Hess 1993).

But there *is* a problem with Fujisaki's model, and it has to do with the intended relation between the phrase component and the prosodic structure of an utterance. Phrase commands are supposed to reflect phrasing in some way – say, a phrase command at the beginning of each intonational phrase. Empirically, however, this condition cannot always be met. In some cases, the result of a phrase command where it would be expected yields a contour of the wrong shape; to get the right shape, we have to put phrase commands in places that make no sense linguistically. Obviously, the appropriate goal of Fujisaki's model (or any model) is not just to model the F_0 contour as a physical signal, but to be able to relate the location of significant events in the contour to linguistic variables. Möbius (1993) has discussed various problems in providing plausible linguistic motivation for the location of certain phrase commands that are needed to make the contour come out the right shape.

A related problem with Fujisaki's model bears on its claim to be generalisable beyond Japanese. Both Liberman and Pierrehumbert (1984) and Taylor (1995, 2000) have attempted to apply Fujisaki's model to English, but have been unable to reproduce certain English contour types – without, again, arbitrarily locating certain phrase commands and/or arbitrarily specifying the size of the phrase commands. The principal difficulty is in modelling low or low-rising accentual contours, such as the one in (2) – a feature that is completely absent from Japanese.

(2) 

Good morning.

The quantitative details of Fujisaki's model are such that negative accent commands yield contours of the wrong shape. It is possible to approximate the low-rising contours by negative phrase components, but this is inconsistent with the intended function of the phrase component.

Finally, even if we concentrate on contour shapes that approximate those for which the Fujisaki phrase component was designed, it is doubtful whether a single phrase-level function is adequate for modelling all utterance types. Grønnum, for example, makes a strong empirical case for meaningful variation of overall slope across a phrase contour in Danish (see figure 1.5). She shows that the steepest slope, falling to the bottom of the speaker's range, is used with complete and final declarative utterances; less steep slopes are used with incomplete declarative utterances and with questions of various types, with the shallowest slope being reserved for echo questions and other questions without morphosyntactic interrogative cues (Thorsen 1980, 1983). On this basis she argues that global slope is a distinctive element of an adequate model of intonation. It is not clear how findings of this general sort could be made compatible with the Fujisaki model.

The most obvious approach to taking account of findings like Grønnum's in a quantitative model of F_0 is to abandon the idea that the phrase component is a single continuous function (like the impulse response of the Fujisaki model), and to modify the realisation of individual accents *locally* by manipulating the value of one or more pitch range parameters. This is essentially the approach taken in the recent mathematical models of F_0 proposed by Kochanski and Shih and their colleagues (e.g. Kochanski and Shih 2003; Kochanski, Shih and Jing 2003), and by Xu and his colleagues (e.g. Xu and Wang 2001; Xu 2005; Xu and Xu 2005). In both cases, the original impetus for these models was to describe the interaction between lexical and sentence-level F_0 factors in Chinese; and in both cases a great deal of attention has been paid to the problems of microprosody and tonal coarticulation (the fine phonetic details of the way one tone influences neighbouring tones). Neither model deals in terms of the impressionistic grid shapes posited by Gårding, yet both can generate a variety of overall shapes by using a number of parameters that influence the realisation of syllable-level F_0 features. The values of these parameters are not constrained to follow a single course specified by a simple function, but can change from syllable to syllable (e.g. pitch range can become wider or narrower to reflect

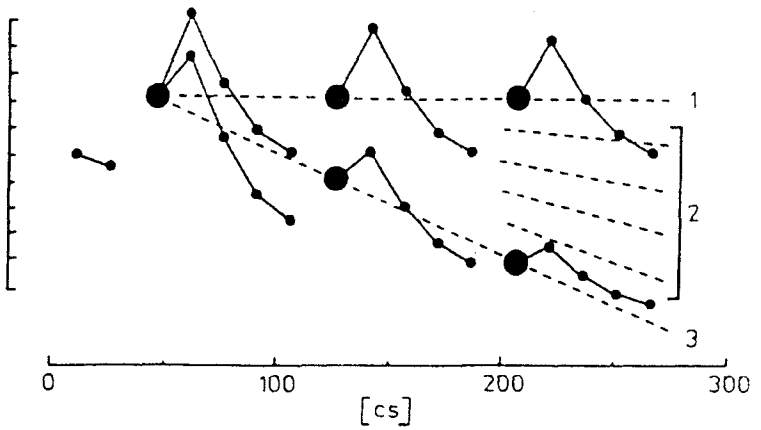


Figure 1.5. *Grønnum's model of Danish sentence intonation. The sentence intonation component is modelled as a straight line of variable slope: nearly level for intonationally marked interrogative (line 1 on the figure); steeply declining for completed final declarative (line 3); intermediate for other interrogatives and for incomplete declaratives (four lines labelled 2). The local low–rise–fall accentual configurations are fitted to the sentence intonation line. From Thorsen 1983.*

differences of 'prominence') or at specified points in the utterance (e.g. pitch range can become narrower following the focused word of an utterance).

Even more than Fujisaki's model, quantitative models such as Kochanski and Shih's, or Xu's, appear broadly compatible with a sequential intonational phonology. Ultimately, any phonological model of intonation will have to be accompanied by a detailed quantitative characterisation of how the individual phonological events are realised phonetically, and how they interact with one another; and models like Xu's or Kochanski and Shih's, in which realisation parameters can be modified in a variety of ways, can easily be used to do just that. That is, the issue of sequential and overlay models discussed in this section may have already turned into the more tractable question of how F_0 realisation parameters interact. However, I have included these models in the discussion here because they are still, to a considerable extent, based on the assumption that intonational meanings or functions (such as focus and interrogativity) are directly signalled by the variability of acoustic parameters, and not mediated by the occurrence of phonological events. In other words, the more basic question of how intonation conveys meaning discussed in section

1.3.1 remains a fundamental point of disagreement. Xu's work, in particular, is clearly motivated by a purely parametric approach to intonational meaning (see especially Xu and Xu 2005). It seems unlikely that those who find the purely parametric approach attractive will take a few pieces of evidence for phonological structure as signalling the end of the debate.

1.3.3 *Intonation and lexical pitch features*

The overlay approach is based not only on Bolinger's wave metaphor quoted earlier, but on a widespread (and partially justifiable) view of the relation between the postlexical features of intonation and lexical features such as tone or word accent. A traditional way of thinking about this relation is that all languages have 'intonation', and that, in addition, some languages have local pitch perturbations for word accent or tone overlaid on the global intonation. There is, as we will see at several points throughout the book, some phonetic basis for this idea; and in fact I believe it is an important challenge for sequential phonological theories of intonation to deal with tone–intonation relations of this sort. It is no accident that recent parameter-interaction approaches, such as Kochanski and Shih's or Xu's, were motivated initially by attempts to model F_0 in sentence context in Chinese, where lexical–postlexical interactions are a central empirical issue.

Nevertheless, I believe that even in its modern incarnation, the overlay view is not justified in the corollary it draws for the relation between intonational function and intonational phonetics. Broadly speaking, traditional overlay models assume not only that all pitch features have duration and extend over domains of different sizes (such as the syllable or the phrase), but also that the domain to which a pitch feature applies functionally must be the domain over which it extends phonetically. In one common formulation (e.g. Pike 1948: ch.1; Laver 1994: ch.15), 'tone' is a feature of syllables; 'pitch accent' ('word pitch', 'word tone', etc.) is a feature of words; and 'intonation' is a feature of phrases or sentences. That is, the traditional view assumes that *the function of a pitch feature should be reflected in its phonetic extent.* But there is empirical evidence that this is not true, and that pitch features whose function is clearly 'intonational' can nevertheless be localised quite precisely in the utterance. This was first clearly demonstrated in Gösta Bruce's work on Swedish word accent (Bruce 1977). Bruce's specific concern was to develop an account of how lexical accent distinctions in Stockholm Swedish are manifested phonetically in different sentence contexts; but his solution to this problem lays the

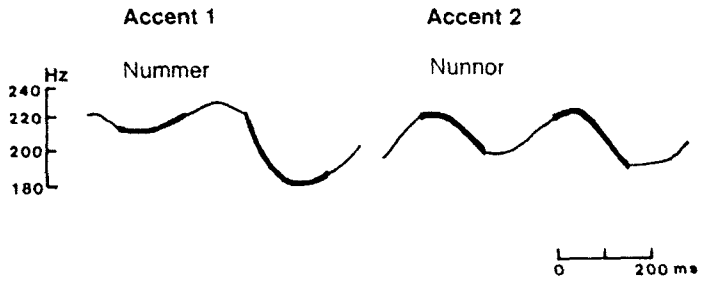


Figure 1.6. *The Swedish word-accent distinction. Citation form contours are shown for the words* nummer *('number', Accent 1) and* nunnor *('nuns', Accent 2). From Bruce 1977.*

foundation for a more general theory of how word-level and sentence-level features interact.

Swedish, like Norwegian and a handful of other European languages, has a morphologically and lexically conditioned distinction of *accent type*. The main stressed syllable of each word, in addition to being stressed, bears one of two accents, called acute and grave, or simply Accent 1 and Accent 2. The classic Swedish minimal pair is *anden* 'the duck' (Accent 1, morphologically *and* + *-en*) and *anden* 'the spirit' (Accent 2, morphologically *ande* + *-en*).[6] The phonetic difference between the two accents is very striking in some environments and exceedingly subtle in others, but it typically involves a difference in the pitch contours of words, and is therefore often described as a difference of 'pitch accent' or 'word accent'. Figure 1.6 shows pitch traces for the 'citation forms' (the forms in which words are typically spoken in isolation) of two segmentally similar words that differ in accent type.

As can be seen from figure 1.6, the phonetic difference between the two accent types is superficially a difference between single-peaked (Accent 1) and double-peaked (Accent 2) pitch contours that span the whole word. This is generally true of citation forms in Stockholm Swedish. However, Bruce established that the genuinely distinctive feature for the two accent types is the *alignment of an underlying pitch peak with the stressed vowel*. In Accent 1, this word-accent peak precedes the onset of the stressed vowel by a considerable extent, so that if there are no preceding unstressed syllables the word-accent

6 Since the distribution of the two accents is almost entirely predictable on the basis of phonological and morphological properties of words, there are not many minimal pairs, and most, like *anden/anden*, involve segmental sequences that can be analysed morphologically in more than one way. This is true in the other European 'pitch accent' languages as well.

peak has no overt manifestation; the stressed vowel simply begins low, and the pitch contour across the stressed vowel actually rises. In Accent 2, on the other hand, the word-accent peak and the subsequent drop in pitch more or less coincide in time with the stressed vowel, and are therefore always present in the phonetic F_0 contour.

The superficial distinction between single- and double-peaked word contours actually results from the interaction of word-accent features and sentence-intonation features in citation forms. There are two relevant features of sentence intonation: a peak and a final low. The peak, or 'phrase accent', is theoretically aligned late in the prominent word, while the final low accompanies the end of the utterance; but of course in a citation form both the peak and the final low occur on the same word as the accentual fall. When these intonational features occur after Accent 2, in which the word accent has already produced a clear peak on the accented vowel, the result is a 'second' peak. But when they follow Accent 1, in which the word-accent peak may be phonetically absent, the result is a phonetic rise across the stressed vowel to the single peak in the utterance. The fact that the distinction is manifested in citation forms as a distinction between single peak and double peak is thus essentially a fact about sentence intonation, because citation forms of words are also complete utterances, and hence contain sentence-level intonational features as well. In contexts other than citation forms, as Bruce amply proved, the accentual distinction based on early and late alignment of the fall is invariably present, whereas the intonational features occur only on certain words in a sentence, and therefore do not always influence the F_0 contour associated with the two accent types.

In addition to providing an elegant and convincing solution for a long-standing problem of Scandinavian phonology, Bruce's analysis provides clear evidence of a case in which pitch features whose *function* relates to domains of different sizes interact as a *sequence of phonetic events* rather than by overlaying small-domain features on large-domain ones. That is, the word accents have a lexical function, while the intonational features like the final peak and fall have a sentence-level function; but the word-accent features are linked to stressed syllables, and the intonational features to final syllables, and together they create a single sequence of tonal events that is realised as a single utterance contour. This does not, of course, establish that phonological structure more in keeping with the wave metaphor is impossible, but it strengthens the argument for a rigorously sequential phonological model, because it shows that such a model provides an accurate account of a case that prima facie might be expected to support the overlay approach.

1.4 Paralanguage and intonation

In the preceding section I have tried to show that traditional 'instrumental' approaches to studying intonation are based on assumptions of questionable validity. I have also argued that one can recognise the existence of phonological structure in intonation – in the manner of the 'impressionistic' descriptions of the 1940s and 1950s – and nevertheless produce methodologically sound work that accounts for the same general range of instrumental data. However, there is still certainly a problem in determining the phonological structure of intonation. The problem arises from the close link between intonation and 'paralinguistic' features in speech.

1.4.1 Relation between linguistic and paralinguistic features

The terms *paralanguage* and *paralinguistic* appear to have been coined in the 1950s to cover aspects of vocal communication that are clearly meaningful but not apparently organised along linguistic lines. (The basic reference is Trager 1958; for a good discussion of the history of the terms and the related ideas see Crystal 1969: ch. 2.) Paralinguistic messages deal primarily with basic aspects of interpersonal interaction – such as aggression, appeasement, solidarity, condescension – and with the speaker's current emotional state – such as fear, surprise, anger, joy, boredom. They are non-propositional and difficult to paraphrase precisely, and yet in many circumstances they communicate powerfully and effectively. Paralinguistic cues can be consistently interpreted even in the absence of the linguistic message – for example, paralinguistic meaning often comes through when the linguistic message is experimentally obscured by such means as acoustic filtering; when the linguistic message is in a language that the listener does not understand; and in some cases even when the listener is non-human. That is, stretches of speech can be produced in such a way as to convey, *irrespective of the linguistic message*, that the speaker is relaxed or impatient or aggressive or whatever.

The relation between paralinguistic and linguistic features is, by and large, unproblematic. The paralinguistic signals represent a parallel channel, or channels, of information, and do not for the most part alter or obscure the identity of the linguistic elements. Someone may call my name angrily, or hesitantly, or with surprise, but it is still unambiguously my name. Nevertheless, a few comments about the relation between the two types of information are in order.

First, the parallel paralinguistic channels are often tightly coordinated in time with the linguistic channel. Much research has shown that nods, hand gestures, and eye contact coincide very precisely with events in the spoken message (e.g. McNeill 1992; Kendon 2004): for example, hand movements often coincide with stressed syllables, and gestures frequently accompany, say, the introduction of a new entity into the discourse (Levy and McNeil 1992).

Second, the paralinguistic message obviously affects the interpretation of the utterance as a whole. Pike's formulation from the 1940s can scarcely be improved on:

> the hearer is frequently more interested in the speaker's attitude than in his words – that is, whether a sentence is 'spoken with a smile' or with a sneer ... If one says something insulting, but smiles in face and voice, the utterance may be a great compliment; but if one says something very complimentary, but with an intonation of contempt, the result is an insult. A highly forceful or exciting statement in a very matter-of-fact intonation may, by its lack of balance, produce one type of irony. (1945: 22)

In interpreting any utterance, in other words, we take into account all the communicative channels, not just the linguistic one. But that fact does not mean that the identity of the channels is confounded; on the contrary, Pike's reference to mismatches between the linguistic message and the attitude with which it is conveyed suggests that the two remain clearly distinct.

Finally and most importantly, we need to mention the nature of the phonetic interaction between linguistic and vocal paralinguistic cues. While I have just argued that the identity of the linguistic and paralinguistic is not normally confounded, it is certainly true that some paralinguistic messages are carried on the same acoustic signal as the linguistic message. How do the two interact?

I believe that, at a fairly low level of analysis, paralinguistic cues should be regarded as *modifications of the way in which phonological categories are realised*. (As I noted in section 1.3.2, Xu's and Kochanski and Shih's models of the interaction between tone and postlexical properties in Chinese can be seen as instantiating exactly this idea.) Such modifications do not normally affect the categorical nature of the phonology; for example, it is known (Tartter 1980) that smiling has an effect on vowel formants, and that this effect can be detected by listeners in recorded speech (i.e. without seeing the speaker's face). The paralinguistic message modifies the acoustic realisation of the vowel categories, but does not modify them so much as to distort their identity. Closer to the topic of this book, we may note that paralinguistic modifications of pitch range – such as raising the voice in anger – modify the realisation of lexical tones in languages like Yoruba or Chinese, but do not normally obscure

their identity. This is true even in languages in which the tone phonemes are distinctive levels (such as High, Mid, and Low in Yoruba); in some way that we do not entirely understand, the phonological essence of the levels must be invariant relative to a phonetic frame of reference that can be modified for paralinguistic purposes.

Distortion of phonological identity is not impossible, however. In some cases paralinguistic modification may cause instances of a given phonological category to sound like instances of some other. A clear example of this is provided by a paralinguistic lip-pursing gesture which is very characteristic of German speech; it seems to convey something like friendliness or reassurance, or some other kind of empathy with the hearer. The acoustic effects of this gesture are, not surprisingly, rather similar to those of phonologically specified lip rounding, and a phonologically unrounded vowel spoken with paralinguistic lip pursing may sound more like a rounded vowel than like a typical unrounded one. This is sometimes represented orthographically in written dialogue by using the letters for rounded vowels. In German comic books I have observed both *nöö* and *mööönsch* as representations of pursed-lips pronunciations of, respectively, *nee* ('no', colloquial) and *Mensch* ('man!', colloquial interjection). In Kurt Vonnegut's World War II novel *Slaughterhouse-Five*, German railway guards are reported as lazily saying 'Yo, yo' (presumably *Ja, ja* 'Yes, yes' with paralinguistic lip pursing) to American prisoners of war who have just reported that one of their fellows has died in transit.

Nevertheless, such cases are relatively rare, and the only people who seem to notice the phonetic 'neutralisation' involved are non-native speakers, and dialogue writers who have the freedom from orthographic convention allowed by the comic book medium. Normal listeners hear the phonological categories and the phonetic modifications in their separate channels. Moreover, languages seem not to allow many such ambiguities to arise – for example, Yoruba in general seems not to use paralinguistic modifications of pitch range to signal greater emphasis on a particular word or phrase (see Rowlands 1969: 24ff.), presumably because this would result in too many occasions where lexical identity could be obscured or neutralised.

1.4.2 *Defining properties of paralanguage*

While the distinction between linguistic and paralinguistic as just described seems fairly clear, there is one area where it breaks down quite substantially. This is in the area of intonation. The most obvious paralinguistic cues are global properties of the speech signal, such as loudness, voice quality, and pitch range. The properties involved in intonation are therefore also involved in conveying

paralinguistic messages. This means that in analysing intonation we are faced with the task of distinguishing linguistic and paralinguistic aspects of the same signal. In principle, as we just saw, this is not a problem; but in practice, in a language like English, it can be very tricky to separate paralinguistic effects on F_0 from what I am calling intonational effects. It is therefore important to try to understand the difference between paralinguistic and linguistic signalling before we tackle intonational phonology.

The central difference between paralinguistic and linguistic messages resides in the *quantal* or *categorical* structure of linguistic signalling, and the *scalar* or *gradient* nature of paralanguage. In linguistic signalling, physical continua are partitioned into categories, so that close similarity of phonetic form is generally of no relevance for meaning: that is, /θ/ and /f/ are different phonemes in most varieties of English, despite their close phonetic similarity; and pairs of words like *thin* and *fin* are not only clearly distinct but also semantically unrelated. In paralinguistic signalling, by contrast, semantic continua are matched by phonetic ones. If raising the voice can be used to signal anger or surprise, raising the voice a lot can signal violent anger or great surprise. Paralinguistic signals that are phonetically similar generally mean similar things. This feature of paralinguistic signalling is often referred to as 'gradience' (Bolinger 1961a); in my view it is the defining characteristic. The difference between language and paralanguage is a matter of *the way the sound–meaning relation is structured.*

I emphasise this in order to draw attention to three aspects of paralanguage which may at first appear to be defining characteristics, but on closer inspection are not. First, paralanguage is not a matter of specific acoustic properties. Voice quality is commonly, perhaps even universally, used paralinguistically, but there are numerous examples of lexical 'phonemic' distinctions of voice quality as well. For example, in Dinka, a major Nilotic language of Southern Sudan, there is a systematic distinction of voice quality running through the lexicon, such that most vowel phonemes can occur with both 'breathy' and 'non-breathy' voice quality; there are many minimal pairs, such as [dɪ̤t] 'big' versus [dít] 'bird'. In many varieties of Chinese, some tone phonemes are distinguished not only by characteristic F_0 patterns, but by differences of voice quality as well (e.g. Rose 1989, 1990). And contrasts based largely on voice quality (often referred to in the relevant literature as 'register', following Henderson 1952) also play an important phonological role in many languages of South East Asia; as Ladefoged humorously noted in a slightly different context (1983: 351), 'one person's voice disorder is another person's phoneme'. There is nothing exclusively paralinguistic about voice quality, and *mutatis mutandis*, the same applies to other common paralinguistic cues as well.

Nor is the difference between language and paralanguage a matter of the domain over which a particular acoustic property is present. Some paralinguistic features are indeed a matter of what Laver (1980) has called 'long-term settings', such as an overall raising of the voice to express active emotions like anger or surprise. But it is important not to equate 'paralinguistic' with 'long-term': paralinguistic features may also be linked phonetically to individual parts of the message, and their effect localised accordingly. Indeed, by coupling paralinguistic cues to individual words or phrases, we can achieve a level of expressiveness that transcends the signalling power of either words or paralanguage functioning separately. (Imagine, for example, an unsympathetic lecturer attaching paralinguistic cues for sarcasm to a single word repeated from a naive question by a student.) Crystal's transcriptions of English natural speech (in e.g. Crystal 1969) contain many indications of paralinguistic markers attached to single words or phrases.

Finally, the difference between linguistic and paralinguistic signalling is not simply a matter of the kind of message conveyed. Obviously, it is impossible to express propositions like '$E = mc^2$' or 'Sardinia is the second largest island in the Mediterranean' without resorting to true language, but it does not follow that it is impossible to use true language to convey subtle emotional or interpersonal messages. Probably all languages contain lexical items and grammatical categories whose main function is to convey nuances of doubt, irony, surprise, condescension, solidarity, and the like. These include categories of modality in verb inflection (e.g. the choice between the subjunctive and indicative in relative clauses in French or Italian); derivational categories like diminutives, augmentatives, and pejoratives (in Italian and many other languages); and 'particles' like German *doch, wohl, eben, ja*, and the like. These can be every bit as effective as strictly paralinguistic cues for conveying emotional or interpersonal meanings.

On all three of the criteria just discussed – its acoustic nature, the domain over which it extends, and the messages it conveys – intonation clearly 'feels' paralinguistic. First, the suprasegmental features that are used in intonation are commonly used in paralinguistic signalling: overall pitch range, loudness, and duration are almost certainly used paralinguistically in all languages. Second, at least some intonational features seem to spread over longer stretches of speech, and it is tempting to view the pitch contour of a sentence as a 'long-term setting' for the sentence as a whole. Third, intonational meaning is unquestionably affective and interpersonal, and almost never propositional: there are not many cases where intonation can reasonably be said to affect the truth value of an utterance.[7] But the claim made here is that none of these three factors is decisive. The real issue is whether intonation involves a categorical structure.

1.4.3 *Intonational meaning*

Let us return to the issue of 'instrumental' and 'impressionistic' approaches to intonation. With the foregoing discussion behind us, it is now possible to see more clearly what the problems with these two approaches are. The difficulty with the 'instrumental' tradition is straightforward: traditional instrumental research of the sort discussed above *treats all intonational meaning as paralinguistic*. It looks for continuously variable phonetic properties of utterances that can be directly related to aspects of the utterance's meaning. While this approach is clearly appropriate for truly paralinguistic meaning (such as overall raising of the voice to convey greater emotional arousal or interest), it is clearly inappropriate for linguistic meaning (as the very idea of 'acoustic cues to negation' or 'acoustic cues to tense' makes clear). By approaching all of intonational meaning in this way, the traditional instrumental approach rules out a priori the possibility of finding evidence for phonological structure in intonation.

The problem with the 'impressionistic' tradition is more subtle. This approach starts from the assumption that there is phonological structure in intonation. The most basic task of any phonological analysis, as I suggested above, is to identify the categorical or quantal elements in a phonological system and to account for the ways in which the realisation of these elements varies. If intonational and paralinguistic messages are indeed distinct, then one of the sources of variation in the realisation of intonational categories is paralinguistic modification, and the basic task of analysing intonational phonology is to tell intonation and paralanguage apart.

7 The only clear exception to this statement – that is, the only clear cases where intonation does affect truth value – involves accent and the scope of quantifiers. For example, the following pair of sentences (from Vallduví 1990 [1992: 142]) will be true or false under different conditions, depending on what the speaker sprinkled where:

(a) I only sprinkled SALT in the stew.
(b) I only sprinkled salt in the STEW.

The following pair (from Rooth 1985: 164) is comparable:

(a) In Saint Petersburg, officers always escort BALLERINAS.
(b) In Saint Petersburg, OFFICERS always escort ballerinas.

In certain cases, there is broad agreement about which is which. For example, in an accent contour that rises to a peak and then falls to the bottom of the range, it is widely agreed that the height of the peak can vary paralinguistically to convey greater emphasis without affecting the linguistic identity of the contour; that is, the difference between the following two utterances:

(3)

(a) He didn't. (b) He didn't.

would generally be treated as paralinguistic only – a matter of greater emphasis in the second case. But in many cases one description presents as an intonational contrast what another treats as the result of paralinguistic variation. To continue with the example, in some descriptions (e.g. Schubiger 1958; O'Connor and Arnold 1973) there is said to be an intonational difference between a 'high fall' and a 'low fall', as in the following pair:

(4)

(a) He didn't. (b) He didn't.

In other analyses (e.g. Palmer 1922; Crystal 1969) these are treated as paralinguistic variants of the same basic intonational category 'fall'. The basic observations about the different meanings conveyed by the two – the first being more emphatic or more argumentative, the second more understated or more final – could be seen as consistent with either analysis.

Since it is genuinely difficult to tell paralinguistic and intonational messages apart, and since there is no obvious independent body of theory, nor standard of evidence, to which one might appeal in order to decide issues like this, such disagreements have tended to remain unresolved. I have elsewhere (Ladd 1993a) referred to these as 'paralinguistic stalemates'. We may also identify the 'paralinguistic gambit', in which one description is presented as superior to some other on the grounds that it correctly distinguishes the intonational contrasts from the paralinguistic variation, while the rival description has been misled into attempting categorical descriptions of things that are really paralinguistic, or vice versa. Good examples of the paralinguistic gambit include: Lieberman (1967: 175), criticising the British 'nuclear tone' tradition; Bolinger (1951), criticising the American 'pitch phoneme' tradition; Crystal (1969: ch. 4), criticising much previous work, but in particular distancing himself from the rest of the British nuclear tone tradition; Beckman and Pierrehumbert (1986: 307), criticising Ladd (1983a) and his proposal for a feature (raised

peak) in English; and Xu's critiques (especially Xu and Xu 2005) of much previous linguistic work on focus.

Despite these problems, we should not overlook the extent of the common ground in adherents of phonological approaches to intonation. In particular, I think it is legitimate to identify something that we might call the 'Linguist's Theory of Intonational Meaning' (see Ladd 1987a: 638). The central idea of this view is that *the elements of intonation have morpheme-like meaning*. These meanings are very general, but they are part of a system with a rich interpretative pragmatics, which gives rise to very specific and often quite vivid nuances in specific contexts. This view contrasts sharply with the assumptions underlying the instrumental approach, in which it is generally assumed that quite specific meanings, such as interrogation, anger, and incompleteness, are conveyed by rather general phonetic properties, such as overall raising of pitch; and that context-dependent pragmatic inference plays little role in the interpretation of intonational features.

The Linguist's Theory of Intonational Meaning is extremely widespread among linguists of otherwise diverse outlooks. Bolinger and Gussenhoven, for example, have extensively developed the idea of 'intonational morphemes'; much of their work on intonational meaning (especially Gussenhoven 1984; Bolinger 1986, 1989) consists of detailed explications of specific intonational nuances on the basis of postulated general meanings for specific elements of intonational analysis. David Brazil (e.g. Brazil, Coulthard, and Johns 1980), working in the context of mainstream applied linguistics and discourse analysis, developed some interesting ideas on the discourse function of the various nuclear tones of the British tradition which influenced Gussenhoven's early work. Liberman (1975: sect. 3.2.5) discusses the general properties of intonational meaning at some length, emphasising the generality of the basic meanings and the vividness of specific nuances in specific contexts. My own work on 'stylised intonation' (Ladd 1978) is a concrete application of this general point of view; the same can be said of the work of Pierrehumbert and Hirschberg (1990) and Steedman (1991, 2000), who attempt to provide basic meanings for the elements of Pierrehumbert's description of English in order to explain the pragmatic implicatures produced by specific combinations of tune and text.

In short, linguists may have markedly different views about what the phonological categories of intonation are, but by and large they agree on how those categories contribute to the meaning of an utterance. The fact that this common understanding has not led to detailed agreement about the analysis of intonation does not necessarily mean that the Linguist's Theory is wrong, but can be taken

as an indication of the difficulty of unravelling intonation from its paralinguistic context.

1.4.4 Prospect

The close acoustic and semiotic connection between intonation and paralinguistic cues is unquestionably the most important conceptual problem in studying intonation. My goal in this chapter has been, in effect, to show that this is only a problem, not an insurmountable hurdle. I have argued that there are good scientific reasons for trying to draw a clear distinction between paralinguistic uses of suprasegmental features and intonation as I have defined it here. More specifically, I have suggested that intonation has a categorical linguistic structure, consisting of a sequence of phonological events occurring at well-defined points in the utterance, and that if we hope to understand the ways in which gradient modifications of phonetic parameters affect the communicative force of utterances, we must take the phonological structure of intonation into consideration as well. The remainder of the book is devoted to exploring the consequences of that view.

2 *Fundamental concepts of the autosegmental–metrical theory*

This chapter sets out several basic ideas of what I will refer to as the autosegmental–metrical (AM) theory of intonational phonology. The somewhat cumbersome name reflects the intellectual heritage of the theory in American non-linear 'autosegmental' and 'metrical' phonology of the 1970s, but it also embodies certain ideas about intonational structure that I will develop in chapters 7 and 8. The general approach has its origins in three influential PhD theses, namely Liberman (1975), Bruce (1977), and – especially – Pierrehumbert (1980). During the 1980s the theory was applied to a variety of European languages (e.g. Ladd 1983a; Gussenhoven 1984) and served as the basis of a number of systems for synthesising intonation by rule (e.g. Pierrehumbert 1981; Anderson, Pierrehumbert, and Liberman 1984; Ladd 1987b). It was central in the development of 'laboratory phonology' (Beckman and Kingston 1990; Pierrehumbert, Beckman and Ladd 2000), in that it very early adopted the goal of explaining details of instrumentally measured F_0 *in a phonological description*; it takes for granted that instrumental phonetics is a source of data for phonological theory. Since the early 1990s, it has given rise to a whole series of ToBI (Tone and Break Index) transcription systems for the intonation of a variety of languages, and on this basis it is widely presupposed in discussions of prosodic structure and of the relation between intonation and information structure.

Like the IPO theory, which was used in chapter 1 to illustrate the assumptions of a 'phonological' but still phonetically accountable approach to intonation, the AM theory adopts the phonological goal of being able to characterise contours adequately in terms of a string of categorically distinct elements, and the phonetic goal of providing a mapping from phonological elements to continuous acoustic parameters. However, unlike the IPO theory, which has its roots in work on speech perception, the AM theory grows out of theoretical problems in phonology – in particular, the general questions of suprasegmental structure that motivated early autosegmental work on African tone systems (e.g. Leben 1973; Goldsmith 1976) and the specific issue of whether to describe

intonation contours in terms of 'levels' or 'configurations' (Bolinger 1951; Smith 1955). As a result there are certain differences of emphasis, and several matters on which the two views substantially disagree. Moreover, even within the broad framework provided by the AM theory there are a number of issues of descriptive detail on which researchers differ, and it is important to emphasise the general unity of approach behind the specific disagreements.

This chapter therefore sets out four basic tenets of the AM approach to intonation, comparing them where appropriate to divergent ideas from IPO and other theories, and pointing to differences of detail within the overall AM view. The four can be stated as follows:

(1) *Sequential tonal structure*: tonal structure consists of a string of local *events* associated with certain points in the segmental string. Between such events the pitch contour is phonologically unspecified and can be described in terms of *transitions* from one event to the next. In languages like English, the most important events of the tonal string are *pitch accents*, which are associated with prominent syllables in the segmental string, and *edge tones*, which are associated with the edges of intonational tunes at major prosodic boundaries.

(2) *Distinction between pitch accent and stress*: pitch accents, in languages that have them, may serve as concrete perceptual cues to stress or prominence. However, they are in the first instance *intonational* features, which are *associated with* certain syllables in accordance with various principles of prosodic organisation. The perceived prominence of accented syllables is, at least in some languages, a matter of *metrical strength* and/or *dynamic stress*, which can be distinguished from pitch accent.

(3) *Analysis of pitch accents in terms of level tones*: pitch accents and edge tones in intonational languages can be analysed as consisting of primitive *level tones* or pitch targets, High (H) and Low (L).

(4) *Local sources for global trends*: the phonetic realisation or *scaling* of any given H or L tone depends on a variety of factors (degree of emphasis, position in utterance, etc.) that are essentially orthogonal to its identity as H or L. Overall trends in pitch contours (e.g. gradual lowering of overall range) mostly reflect the operation of *localised* but *iterated* changes in scaling factors.

The chapter concentrates on the empirical motivation for these four tenets, on showing how they contribute to a coherent theoretical whole. The discussion is, from the point of view of an adherent of Pierrehumbert's work on English or of the ToBI movement, deliberately elementary. This style of presentation has a twofold goal: first, I aim to defend the AM approach in terms that will be accessible to anyone concerned with intonation, not just to those who are already familiar with issues in theoretical phonology or the transcription of specific languages; second, I wish to identify what I see as the essential ideas of the approach, and to distinguish them from the details of specific analyses of specific intonation systems. Closer consideration of Pierrehumbert's description of English and of ToBI transcription systems for other languages will be undertaken in chapter 3.

2.1 Sequential tonal structure

As suggested in chapter 1, an important feature of any phonologically oriented description of intonation is that it represents pitch contours phonologically as sequences of discrete intonational events. For languages like English and Dutch, the AM theory assumes that there are two main types of such events, pitch accents and edge tones. In tone languages and other languages with lexically specified pitch features, tonal events may have different functions, but, as we shall see in section 4.3, the basic phonological structure is essentially the same.

It is not difficult to demonstrate the usefulness of distinguishing between pitch accents and edge tones. Consider, for example, the rising–falling–rising tune that can be used in English for a strongly challenging or contradicting echo question, as in the following exchange:[1]

(1) A: I hear Sue's taking a course to become a driving instructor.

B: Sue!?

In the traditional British analysis, this contour would be described as a 'rise–fall–rise' nuclear tone, and on the basis of the monosyllabic utterance the rising–falling–rising melody appears as the essence of the contour type under consideration. However, we are not dealing with a global rise–fall–rise *shape*

1 In order to appreciate the point of these examples it is important to get the intonation right on B's reply, and the reader is encouraged to listen to the online sound files of this and the other examples in the book. We are not interested in the more or less steadily rising contour that conveys surprise or merely a request for confirmation. The relevant tune is one that, on the monosyllabic utterance *Sue*, rises and falls and rises again.

that applies holistically to an entire utterance. This can be seen when we put the same tune on a longer utterance:

(2) A: I hear Sue's taking a course to become a driving instructor.

B: A driving instructor!?

The rise–fall–rise shape that spanned the entire one-syllable utterance in *Sue!?* is not simply stretched out over the six-syllable utterance here. Instead, the contour is seen to consist of a sequence of at least two discrete events, an accentual feature consisting of a rise through a prominent syllable (here *driv-*) followed by a fall, and an edge tone consisting of a rise during the last few tenths of a second of the utterance. The level low stretch on the syllables *-ing instruct-* is simply a transition between these two events, which can be vanishingly short (as on the monosyllable *Sue*) or as long as needed. For the identity of the tune it is the accentual rise-plus-fall and the boundary rise that 'count': they are the distinguishing features that lead us to describe the contour on *Sue!?* and that on *A driving instructor!?* as 'the same'. A description in these terms gives us a simple but accurate way of describing how the tune in question is applied to texts with varying numbers of syllables and different stress patterns.

By breaking down the contour into component parts in this way, we do not, of course, preclude the possibility of referring to larger units. In particular, most of the nuclear tones of the British tradition (e.g. O'Connor and Arnold 1973) can be readily translated into combinations of pitch accents and edge tones (cf. Gussenhoven 1984; Roach 1994). In the example just given, we have not questioned the existence of a rise–fall–rise nuclear tone in English, but have simply been more explicit about its internal structure. The rise–fall–rise nucleus in English can be seen as a composite or superordinate unit, like the 'hat pattern' in the IPO model (cf. section 1.2.1). As we shall see in section 3.1, the basic intonational taxonomy of the British tradition is not for the most part rendered obsolete by the theoretical framework presented here, only cast in a new light. There is no necessary contradiction in recognising both the functional unity of 'nuclear tones' or other such 'tunes', and the phonological separateness of their component parts. On the other hand, by distinguishing accents from edge tones we may be able to resolve some long-standing problems in the traditional British taxonomy. In particular, much of the difficulty with the variety of falling–rising tunes in English (see e.g. Lee 1956; Sharp 1958; O'Connor and Arnold 1973: 29; Gussenhoven 1984: ch. 3) can be seen as a consequence of

failing to make this distinction: there does seem to be a characteristic rise–fall–rise intonational unit in English, but it can probably be distinguished from essentially adventitious sequences of a falling accent and an independently selected boundary rise.

I should note that I am using the term 'edge tone' here as a general term for any tone that is associated with the periphery of a prosodic domain. For languages like English, the term covers Pierrehumbert's 'phrase accent' and 'boundary tone', both of which occur at or near the end of a prosodic phrase, after the last pitch accent. However, other descriptions of various languages have also posited edge tones occurring at the beginning as well as at the end of prosodic domains. Moreover, edge tones do not necessarily have to be manifested phonetically at the very edge of a domain. For example, Pierrehumbert and Beckman (1988: 126ff.) have proposed that the initial rise characteristic of prosodic words in Japanese is due to the presence of an H edge tone that can be aligned with either the first or second mora of the word, and for this reason they propose to treat such tones as primarily associated with nodes in a tree rather than with 'edge' elements in the segmental string. Gussenhoven's analyses of lexical accent distinctions in Limburg Dutch dialects (e.g. Gussenhoven 2000) similarly depend on the idea that edge tones need not occur phonetically at the very edge of a domain. Edge tones are discussed further in sections 3.1.4, 4.1.4, and 4.3.

One key innovation of the AM theory here is that it draws an explicit distinction between *events* and *transitions*. It recognises that certain localised pitch features are linguistically important, while much of the rest of the pitch contour is merely what happens between the important features. This differs from other theoretical conceptions of pitch phonology in two significant ways. First, the AM view assumes that the important features are localised 'events', not long stretches of contour. This is an important point of difference between the AM theory and the IPO theory: the latter allows linguistically specified line segments (the 'Type 4 Rise' and 'Type D Fall') that may span several syllables. From the point of view of the AM theory (as has been pointed out by Ladd 1983a: 747ff. and Cruttenden 1992a), such globally rising or falling line segments can simply be treated as the gradual transition from the pitch level at the end of one local event to the very different pitch level at the beginning of the next.

The other key innovation of the AM approach to describing pitch contours is that it ascribes no necessary role in pitch description to the *syllable*. Traditional narrow phonetic transcriptions of pitch in the British tradition involved representations of the relative pitch of each successive syllable, arranged on a kind

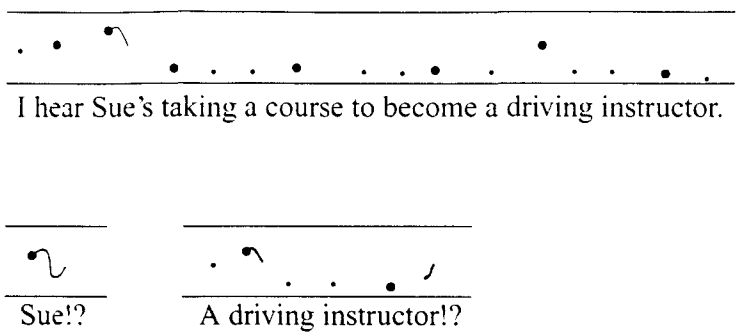


Figure 2.1 *Detailed syllable-by-syllable impressionistic notation of intonation ('tadpole diagram') of the sort commonly used by traditional descriptive works of the British school of intonation analysis.*

of musical stave. Such diagrams (informally known as 'tadpole diagrams' from the shape of the marks used to represent moving pitch on prominent syllables) are illustrated in figure 2.1. Similar diagrams are produced automatically by the Prosogram software developed by Piet Mertens (2002; cf. also D'Alessandro and Mertens 1995; Hermes 2006). While such a syllable-by-syllable transcription is one intuitively natural way of describing the phonetics of pitch, it is by no means clear that the use and perception of intonation involves resolution into syllable pitches. Even early British users of such notation systems (e.g. O'Connor and Arnold 1973: 7–13) pointed out the functional equivalence of a fall in pitch on one syllable and a high–low sequence on two syllables, for example. By treating the fall as an event in its own right, and treating the association between the fall and the accompanying text as a separate issue, the AM view easily expresses this functional equivalence.

2.2 Pitch accents, prominence, and tune–text association

As we just saw, the phonological elements of the pitch contour that accompany certain stressed syllables in languages like English or Dutch are known in the AM theory as pitch accents. A pitch accent may be defined as a local feature of a pitch contour – usually but not invariably a *pitch change*, and often involving a local maximum or minimum – which signals that the syllable with which it is associated is *prominent* in the utterance. Pitch accents have a dual aspect: they are simultaneously building blocks of intonation contours and important cues to the prominence of the syllables with which they are associated. This use of

the term 'pitch accent' was first proposed by Bolinger (1958) and resurrected by Pierrehumbert (1980), and is now in general use. It should not be confused with a competing sense of the term, denoting the *lexically specified pitch features* of languages like Japanese, some Bantu languages, and some European languages, including Swedish, Norwegian, and Slovenian. Where necessary, the two senses can be distinguished as *intonational pitch accent* and *lexical pitch accent*.

Bolinger originally proposed the notion of (intonational) pitch accent as a way of making sense of experimental data (in particular the findings of Fry 1955, 1958) showing that pitch change is a major cue to the perception of stress. In Bolinger's view, 'stress' is nothing but an abstract lexical property of individual syllables, while 'pitch accent' is actual prominence in an utterance. If a word is prominent in a sentence, this prominence is realised as a pitch accent on the 'stressed' syllable of the word. Much the same view was put forth early in the development of the IPO model, where pitch accents were known as 'prominence-lending pitch movements', and little consideration was given to supposed differences of prominence not involving pitch movement. This view is attractive for its simplicity, but there is now overwhelming evidence that it is empirically inadequate. Specifically, as we shall see in the next section, it seems clear that linguistically significant prominence distinctions can be signalled in a variety of ways that do not involve pitch accent. To clarify this issue, we need to take a more detailed look at the whole problem of stress.

2.2.1 *The phonetic nature of stress*

Probably no topic in the general area of intonation and suprasegmentals has posed such a puzzle as stress. A great deal of experimental work going back many decades has sought to establish both physiological and acoustic correlates of stress as a phonetic phenomenon. Lehiste, writing in 1970, says that 'of the three suprasegmental features [quantity, tonal features, and stress], stress has for a long time been the most elusive one' (p. 106).

A simple example makes clear the extent of the problem. Consider a stress minimal pair from English, like *permit* (noun) versus *permit* (verb). In citation form, the stress contrast is signalled most conspicuously by differences in pitch contour:

(3)

The pitch rises to a peak on the stressed syllable, followed by a rapid fall. There are also differences of vowel quality, intensity, and especially syllable duration between the two words, but, as we just saw, Fry (1958) showed that the pitch differences are the cues that listeners use most reliably to make judgements about stress in isolated words when the suprasegmental features are manipulated experimentally. Yet if we change the citation forms into questions, the pitch contours are much less distinct:

(4)

(a) p e r m i t (noun) (b) p e r m i t (verb)

Moreover, the question contours are completely different from the statement contours; we can no longer say that the stressed syllable is cued by a pitch peak. And if we put the words in context after the main intonational peak of an utterance, there may be no pitch distinction at all, yet we still clearly perceive the different stress patterns of the noun and the verb:

(5)

(a) I TOLD you the permit had expired!

(b) I TOLD you they'd permit him to retire!

A long-standing interpretation of Fry's results is that 'stress' – in the sense of perceived prominence on a syllable – is a complex amalgam of F_0, duration, and intensity, with F_0 generally the most important, but with uncomfortable exceptions made for certain intonational contexts such as (4) and (5). This view has remained remarkably widespread in certain quarters since the time of Fry's experiments, and a good deal of instrumental work has concentrated on trying to find better characterisations of the correlation between acoustic cues and perceived prominence (e.g. Gay 1978). From the point of view of the AM approach, though, much of this work is misdirected, and ignores basic theoretical problems. The central such problem is the assumption that 'stress' is a scalar phonetic property of *individual syllables*, and that pitch is one of the components of that scalar phonetic property. Instead of adopting this idea, the AM view assumes that 'stress' is – or is partly based on – an abstract

phonological property of a syllable *within a prosodic structure*, and that 'pitch accent' and 'stress' are therefore not the same thing.

Pierrehumbert expresses the difference between the two views as follows:

> In the wake of Fry's classic study ... the impression grew up that F_0 can be viewed as a transducer of stress: the higher the stress, the higher the F_0 (or the greater the F_0 movement). In the [autosegmental–metrical] framework ... the relation of F_0 to stress is not as direct as this. Rather, a ... given F_0 pattern could be compatible with more than one conclusion about the location of stress, if more than one assumption about where the accent is located was consistent with a well-formed intonational analysis for the contour. (1980: 103)

Halliday comments on the same theoretical issue as follows:

> It is thus a little misleading to ask anyone if he can 'hear the four degrees of stress'. The answer may well be that he can hear, and tell apart, what are being *called* four degrees of stress, but would analyse them as something else; but the question is so framed as to preclude this answer. (1967a: 14n.)

The outlines of a theory in which stress is not treated as a single scale of 'degrees of prominence' applying to individual syllables have been growing clearer for some time. In the earliest versions the distinction is drawn between 'abstract' word stress and 'concrete' sentence stress. Lehiste ascribes this view to Weinreich (1954), and supports this position herself, though apparently with some reluctance:

> It appears probable that word-level stress is in a very real sense an abstract quality: a potential for being stressed. Word-level stress is *the capacity of a syllable within a word to receive sentence-stress when the word is realized as part of the sentence* ... The fact that not all syllables that are perceived as stressed are associated with peaks of subglottal pressure supports the idea that what is realized phonetically is sentence-level stress rather than word-level stress. In other words, our knowledge of the structure of the language informs us which syllables have the potential of being stressed; we 'hear' the underlying phonological form. (1970: 150; emphasis added)

More or less the same view was taken by Abercrombie (e.g. 1991), who distinguished *accent* (word-level abstraction) from *stress* (actual phonetic manifestation in an utterance). This view helps with cases like (5) above, but in some sense still leaves us with the same question: instead of 'What are the acoustic correlates of stress?', we have 'What are the acoustic correlates of actual sentence-level stress?' Further theoretical refinement is needed.

Bolinger's pitch accent theory, first published in 1958, was the first proposal for such a refinement. For Bolinger, as for Lehiste or Abercrombie, lexical stress is a phonological abstraction, and the 'stressed' syllable of a word is simply the place where actual sentence-level prominence occurs, if the word is prominent in a sentence.[2] Where Bolinger differs from Abercrombie, and probably from Lehiste, is in claiming that actual sentence-level prominence is more or less *exclusively* a matter of intonational pitch movement. When there is no intonational pitch movement, there is, Bolinger claims, no prominence and no consistent phonetic correlate of lexical stress. As speakers of the language we think we hear the lexical stress because we recognise the word. Thus in the citation forms in (3) above, the peak and fall are intonational pitch features that attach to the lexically stressed syllable when the word is prominent in the utterance (which it is, more or less by definition, in a citation form). The situation in (4) is similar, in that the lexically stressed syllable attracts the intonational pitch movements, though these movements are very different, for reasons having to do with sentence intonation rather than anything directly to do with stress. In (5), where there is an emphatic stress on *told* and consequently no sentence-level prominence on *permit* at all, the stress difference we hear between the noun and the verb in this context is due entirely to our knowledge of the language, and not to any consistent phonetic cues.

Bolinger's view, which was also espoused by the IPO theory, succeeds in making some sense of the relationship between stress and pitch in languages like English or Dutch. In particular, it helps explain why pitch can be an overwhelmingly reliable cue to stress in some contexts (like (3)) and virtually irrelevant in others (like (5)). However, it does little to explain the existence of phonetic cues to stress other than pitch accent, such as duration and intensity. Nor does it explain the fairly consistent impression on the part of analysts that there are *degrees* of relative stress, although Bolinger does note that vowel reduction could provide the basis for another 'degree of stress' (e.g. 1964: 285; 1972a: 22). That is, syllables, for Bolinger, can have either a full vowel or a reduced vowel: if they have a full vowel they can be abstractly stressed or unstressed in the lexicon; if they are stressed in the lexicon, they may or may not be actually prominent in an utterance. Actual prominence in an utterance is signalled by pitch accent alone.

2 Note that Bolinger's and Abercrombie's use of the terms 'stress' and 'accent' are almost the reverse of each other. The lexical abstraction is 'stress' for Bolinger and 'accent' for Abercrombie; the actual utterance prominence is 'accent' for Bolinger and 'stress' for Abercrombie.

A somewhat more complicated version of the same basic idea, proposed in two variant forms by Halliday (1967a) and Vanderslice and Ladefoged (1972), and given a textbook presentation in Ladefoged 1982, says in effect that there are *two* sentence-level prominence features, not just pitch accent. (This view, like Bolinger's, also allows for vowel reduction, and, like Bolinger's, treats it as an independent phonetic or phonological phenomenon rather than as a 'level of stress'.) In what follows I will use the terms (utterance-level) *stress* and (intonational) *accent* to refer to the two sentence-level features, but it should be kept in mind that the writers I am summarising use different terms. For what I am calling 'stress' and 'accent', Halliday uses 'salience' and 'tonicity' respectively; Vanderslice and Ladefoged use 'accent' and a cover feature [± intonation] respectively; and the Ladefoged textbook uses 'stress' and 'tonic accent'.

Regardless of the terminology used for the two sentence-level prominence features, this view sees both features as either–or categories: syllables are either stressed or not, and if stressed they can be either accented or not. In contrast to Bolinger's view of 'stress', in the two-feature view both features have a phonetic basis; neither is an abstraction. 'Accent' covers some of the cases that Bolinger referred to as accented, but is explicitly linked to the structure of the intonation contour, and only includes what are sometimes called 'primary' or 'nuclear' accents. 'Stress' includes the rest of Bolinger's accents, but also other prominent-sounding syllables: Halliday emphasises rhythmic considerations in his definition of stress, while Ladefoged emphasises physiological facts such as an increase in respiratory energy. All versions of this view involve the claim that 'the intonation system interacts with the accentual [i.e. stress] one so as to account for what some observers have analysed as a difference in degree of stress' (Vanderslice and Ladefoged 1972: 820; see also the quote from Halliday above). In short, this view accepts Bolinger's idea that intonational accent is a distinct phenomenon from stress, but disagrees with Bolinger by assuming that stress has a separate phonetic reality of some sort and is not merely a lexical abstraction. To summarise this view as we just summarised Bolinger's, we would say that syllables can have either a full vowel or a reduced vowel; if they have a full vowel they can be abstractly stressed or unstressed in the lexicon; if they are stressed in the lexicon, they may or may not be actually prominent in an utterance; if they are actually prominent in an utterance, they may or may not be pitch-accented. Actual prominence in an utterance is signalled by a complex of phonetic cues that reflect greater force of articulation and possibly rhythmic regularity. Pitch accent is an additional feature that is part of the intonation system.

The view incorporated into the AM theory takes the idea of a distinction between utterance-level stress and intonational accent one step further: instead of allowing a single binary distinction of 'stress' in actual utterance prominence, it assumes that utterances have a 'stress pattern', which may involve several different degrees of perceived prominence. This stress pattern reflects a set of abstract *prominence relations* between the elements of the utterance. The stress pattern is manifested in a variety of phonetic cues, which are admittedly not well understood. In addition to the stress pattern, there is an intonation pattern for the utterance, which is composed of a string of pitch accents and edge tones, and the pitch accents are 'lined up with the text on the basis of the prominence relations' (Pierrehumbert 1980: 102). General conditions of prosodic well-formedness stipulate that pitch accents must occur with prominent stressed syllables, and the occurrence of a pitch accent therefore serves as a *cue* to the location of prominence, as Fry and others have found. The essential nature of pitch accent, however, is not 'prominence-lending', as suggested by Bolinger and the IPO researchers, but merely 'prominence-cueing' (a term suggested to me by Francis Nolan).

This theoretical view has begun to influence instrumental work on stress, and evidence is accumulating that not only duration and intensity, but also various spectral properties, if measured appropriately, can actually serve as reliable indicators of stress in English. For example, Beckman (1986) argues that the traditional distinction between melodic accent ('pitch accent') and dynamic accent ('stress accent') is valid; she shows experimentally that in Japanese, pitch change is essentially the *only* acoustic cue to accent, whereas in English, duration, intensity, and vowel quality all play a significant role (we return to the difference between melodic and dynamic accent in section 4.3). More recent work by Campbell (e.g. 1993) lends substance to this view by providing a way of quantifying duration that identifies English stressed syllables reliably, even in the absence of any intonational pitch features. A widely cited study on Dutch by Sluijter and van Heuven (1996) explicitly starts from the assumption that accent and stress need to be distinguished, and shows that in cases where stressed syllables have no pitch accent (as in examples like 2.5 above), spectral balance – the relative energy in different parts of the spectrum – can be used as a reliable indicator of whether a syllable is stressed. Campbell and Beckman (1997) tried to replicate Sluijter and van Heuven's study for English and reported that their results were not so clear. However, similar cases were studied by Huss 1978 (albeit with much less sophisticated tools for acoustic measurement), who claimed to have found differences between stressed and unstressed in unaccented contexts. A much more recent study by Beaver *et al.* (2007) reaches similar conclusions, though on the basis of rather different methodology and a rather different theoretical focus.

Although there is still no consensus on the details, then, it seems likely that we may appropriately speak of 'stress' as a phonetic reality independent of pitch accent. This still leaves the question of the relation between the two phenomena, and whether it is appropriate to speak of multiple levels or degrees of stress. These are the topics of the next two subsections.

2.2.2 *Tune–text association*

If, as just suggested, utterance-level prominence patterns ('stress') can be characterised independently of the distribution of pitch accents, then a complete account of intonational phonology within the AM approach will have to incorporate a theory of stress or linguistic prominence. The search for such a theory is the focus of much work within what has come to be known as metrical phonology.

Metrical phonology begins with Liberman's notion that linguistic prominence crucially involves a *relation* between nodes in a binary-branching tree structure (Liberman 1975; Liberman and Prince 1977). According to Liberman and Prince, in any such relation one node is strong and the other weak:

(6)

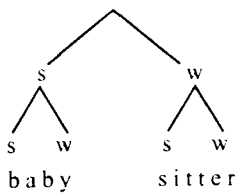
It is important to emphasise (because this is a point on which there has frequently been misunderstanding) that no absolute degree of prominence is implied by the labels 'strong' and 'weak'. There is no direct phonetic interpretation whatsoever of either label, but only of whole structures. What the notation means is that one node is *structurally* stronger than the other; this relative strength may be manifested phonetically in a great variety of ways. This abstract structural understanding of 'strength' – explored in depth by Beckman (1986: chs. 2 and 3) – is crucial for the metrical interpretation of experimental studies of the perception of stress.

For example, consider the two words in (7) in the form they have when pronounced as separate words.

(7)

In citation form, the 'strong' syllables of these two words are phonetically very prominent, with substantial pitch excursions, and the 'weak' syllables are conspicuously less so, with energy dropping off very rapidly and the pitch trailing away to the bottom of the speaker's range. But now let us combine them into a single compound:

(8)

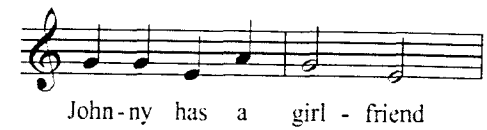


(This representation says that *ba-* is stronger than *-by*, *sit-* is stronger than *-ter*, and *ba-* is stronger than *sit-*. It leaves the relative strength of *-by* and *sit-* unspecified; I will return to the significance of this point below.) The phonetic realisation of this more complex structure, unlike the simple case, allows no single phonetic interpretation of the notion 'strong' syllable. However, it is still possible to state reasonably clearly how the prominence relations are realised. The structure in (8) defines a single strongest point or peak of prominence for the phrase as a whole, on *ba-*. In Liberman and Prince's terminology (1977: 259), this peak of prominence is known as the *Designated Terminal Element* or DTE. Like the citation forms, the compound has a single DTE, and, as in the citation forms, the sole pitch accent occurs at the DTE. This means that in the compound only *ba-* is marked by a pitch accent. We can still identify *sit-* as stronger than *-ter* in the compound, but the phonetic cues are a great deal more subtle, and certainly do not involve any significant pitch change. Pitch change, in other words, serves as a cue to prominence when the structure allows it, but prominence relations can also be cued in other ways. There is thus no problem in principle with describing both *ba-* and *sit-* as stressed syllables in this phrase, even though our understanding of the details of how stress is signalled is still incomplete.

The idea that pitch accents are elements of an intonational tune, and occur at syllables whose prominence in the utterance is somehow independently definable, is an aspect of what Liberman called 'tune–text association'. This notion can be motivated by analogy to the relation between musical tunes and song texts. In setting texts to music, linguistically stressed syllables are not necessarily associated with long or high-pitched or saliently loud notes – as would be expected if stress is primarily a matter of specific acoustic cues. Rather, what is crucial is for linguistically stressed syllables to be associated

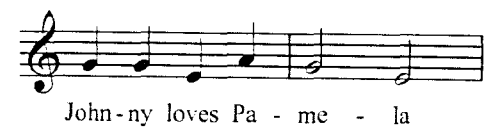
with notes at musically strong positions – 'on the beat', roughly speaking. Liberman's example is the 'children's chant':

(9)



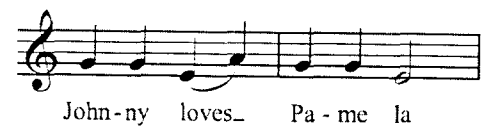
As Liberman points out, ill-formed associations of texts to this tune must be defined in terms of the position of stressed syllables relative to the metrically strong locations in the tune (the notes immediately preceded by bar lines). We cannot chant:

(10)



but must adjust note durations and syllables as follows:

(11)



so that the stressed syllable *Pam-* comes out on the beat. The very fact of occurring on the beat – even though that is in many ways a fairly abstract notion – then serves as a cue to the stress on that syllable.

All of this is at least consistent with the idea that linguistic stress is primarily a matter of phonological strength, and of association with features of the tune; local acoustic properties of individual syllables are not necessarily interpreted locally, but as part of an overarching metrical structure. This general view may seem perversely abstract to those whose interest is in perceptual cues. However, comparable distinctions are well motivated in music theory. In fact, in music we may clearly distinguish between the structural strength or prominence of the 'downbeat' (the main beat of each bar), the actual acoustic prominence of an 'accent' (added loudness, duration, added percussion, etc.), and the occurrence of salient features of the melodic and harmonic structure. For example, some musical styles are characterised by accents (added loudness, etc.) on weak beats (e.g. the 'backbeat' in rock and roll, or the striking off-the-beat accents in reggae). Similarly, in order to define 'syncopation' in music, we must distinguish between the structural strength of the underlying beats and the prominence of melodic events occurring 'off the beat'. Obviously,

the very fact that music theory has terms for things like 'backbeat' and 'syncopation' suggests an expectation that acoustic prominence and prominent melodic events will go together with underlying beats. In speech, too, major pitch movements generally coincide with other cues to prominence. But the importance of distinguishing structural strength from melodic events and acoustic prominence is also clearly shown by the musical analogy.[3]

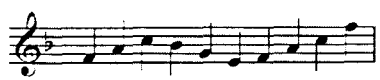
2.2.3 *Metrical strength, dynamic stress, and pitch accent*

We might summarise the foregoing discussion by saying that the 'elusive' nature of stress, on which Lehiste commented, is primarily a theoretical problem and not an empirical one. I believe that metrical phonology, combined with the notion of tune–text association, goes a considerable way to solving this theoretical problem. In particular, it permits us to distinguish three facets of 'stress' that have often been conflated in earlier accounts: abstract prominence relations ('metrical strength'); concrete acoustic prominence or salience on a particular syllable ('dynamic stress'); and the location of prominence-related intonational events ('pitch accent'). However, within metrical phonology itself little effort has been made to account for older instrumental findings about 'stress'. Moreover, some phonologists (e.g. Hayes 1995) have effectively used the abstractness of metrical structure as an excuse not to be concerned with the phonetic manifestation of that structure. Consequently, the AM theory is still not widely accepted by many experimental phoneticians, and the traditional interpretation of Fry's 1958 findings still informs published work (e.g. Levi 2005 on Turkish, which reaches the rather misleading conclusion that Turkish is a 'pitch accent language').

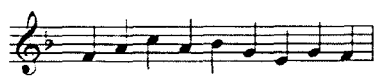
For example, objections have been raised to analysing some kinds of syllables as 'accented' because they are phonetically indistinct or in any case much

3 Note also in this connection that melodic features in music can influence the way in which temporal and other acoustic cues are interpreted: temporally identical sequences of tones may be grouped differently in perception depending on the apparent harmonic progression (Tan, Aiello, and Bever 1981; Dawe, Platt, and Racine 1993; Hannon *et al.* 2004). For example, sequence (a) will be heard with a prominent beat every three notes, while sequence (b) will be heard with a prominent beat every four:

(a)



(b)

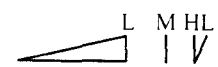


I am grateful to Dik Hermes for the example, to Ani Patel for pointers to the literature, and to Marco Ladd for help in creating the sound files for the two sequences.

less phonetically salient than other syllables in a constituent. A case in point is the utterance-final syllables of French. For many phoneticians, what is striking about these syllables is that they are so often acoustically indistinct, especially compared to the phrase-initial *accents d'insistance* that are prevalent in reading style. Indeed, it has been suggested (e.g. by Beckman 1986: 33) that French does not have stress at all, but only 'demarcative' cues to the edge of a phonological word or phrase; furthermore, there is psycholinguistic evidence that French-speakers are generally insensitive to stress in other languages (Dupoux *et al.* 1997; Dupoux, Peperkamp, and Sebastián Gallés 2001). Yet it has been convincingly argued by Dell (1984) that the distribution of tonal properties of French intonation contours can only be accounted for by assuming that the last full-vowel syllable in a French phonological word has some special structural prominence.

Dell's arguments are based crucially on the tonal behaviour of sentences in which the final schwa or *e muet* is realised. If we assume that accent tones in French go to the edge of the phonological word, then we would expect them to attach to schwa as readily as to any other syllable, but, as Dell shows, this is not what happens. For example, consider the rather emphatic statement intonation that falls (sometimes only slightly) from a high peak on the final syllable of a phrase (the *intonation d'implication* of Delattre 1966). This is illustrated by Dell in the following examples (slightly modified from Dell's originals, but preserving Dell's impressionistic indication of a 'Mid' tone):

(12)



(a) T'étais pas au courant. 'You didn't know what was going on.'

(b) Il met la tab'. 'He's setting the table.'

When this contour is applied to a sentence with any of various kinds of tags – such as vocatives or right-dislocated[4] constituents – the final L spreads to the

4 'Right-dislocation' refers to the occurrence of a constituent (normally a noun phrase) following a clause, with a pronoun or other grammatical marker indicating the grammatical place that the constituent 'would have' occupied had it occurred inside the clause. A possible English example is *She's clever, that lass*, in which *that lass* is right-dislocated and *She* indicates its grammatical role. Outside Celtic-influenced varieties, right-dislocation is not very common in English. It is widespread in all the Romance languages.

whole tag constituent, which forms a separate prosodic phrase (*tronçon*); the H remains anchored to the last syllable of the first prosodic phrase:

(13) L M H L

Il met la tab', Mercier. 'Mercier's setting the table.'

As Dell points out, if the final schwa of *table* is realised in this context, it does not count as the last syllable of the first prosodic phrase, but rather the H is still anchored to the syllable *ta-*:

(14) L M H L

Il met la table, Mercier. 'Mercier's setting the table.'

The last full-vowel syllable has some special status as an anchor for tune–text association; we cannot simply say that the tones go to the edge of the phrase.

Given the metrical conception of prominence under discussion here, Dell identifies the special status of the last full-vowel syllable as 'accent' or 'metrical strength'. In Dell's words,

> the criteria that determine the position of the key syllables [= DTEs] within an intonational phrase are exclusively accentual (and not syntactic). It is true that the relevance of accentual factors ... is not as striking in French as in some other languages. This is because in works on intonation one normally starts by investigating the phrase-final melody that precedes silence. In this context, as a result of certain peculiarities of the phonology of schwa, words are always accented on their last syllable. Consequently, certain melodic events may give the illusion of being associated with the right edge of constituents, whereas in fact they are associated with accented syllables. (Dell 1984: 67f.; my translation)

Dell clearly identifies the prominence of the last full-vowel syllable as structural, not phonetic, based on abstract metrical strength and the association with 'certain melodic events', that is, pitch accents. The fact that this syllable may not bear 'dynamic stress' in Beckman's sense – which is seen as central by many phoneticians – is therefore irrelevant to his analysis. At the same time, the apparent dissociation between dynamic stress and the more abstract phonological characteristics may go some way to explaining what Dupoux and his colleagues call the 'stress deafness' of French-speakers.

In this connection it may also be useful to note the difficulties encountered by Odé (1994) in investigating Indonesian intonation within the IPO framework.

In keeping with IPO assumptions, she attempted to get listeners to identify 'prominent' words, but her subjects – described as 'graduate linguists' – found the task very difficult and disagreed considerably among themselves. Among her conclusions is the claim that 'prominence in Indonesian cannot be described in terms of stressed or accented syllables' (p. 63). Yet acoustic phonetic descriptions of Indonesian intonation in the same volume (Ebing 1994; Laksman 1994) all agree that characteristic pitch movements in Indonesian can readily be defined, mostly with respect to the penultimate syllable of the word. This means that in the terms proposed here, the penultimate syllable in Indonesian is 'accented', because it is the anchor for the events of the intonational string. However, it is not necessarily 'stressed' in the sense of having some clear phonetic salience, because that is a separate phenomenon, which apparently does not occur very consistently in Indonesian.[5]

2.2.4 *Summary*

The foregoing discussion of stress and accent in the AM theory can be summarised as follows. In at least some languages, certainly including English and Dutch, there is a phonetic phenomenon of stress that can usefully be distinguished from pitch accent. Stress in this sense might be glossed as 'acoustic salience': it is a complex of properties that can be related to greater force of articulation, including increased intensity and duration, and shallower spectral tilt.

The prominence pattern of an utterance reflects the organisation of the syllables into a hierarchical metrical structure. This structure specifies abstract relations of prominence or strength between syllables, and between larger constituents such as words and phrases. The way in which words and phrases are fitted into a well-formed metrical structure is dictated in part by the lexically specified stress pattern of words.

By and large, syllables that are prominent in the abstract metrical structure are also phonetically stressed, but the theory does not preclude the occurrence of systematic dissociations. A syllable can be metrically strong or prominent without necessarily being stressed. Such a dissociation is comparable to

5 In this connection, consider also Laksman's discussion of the fact that pitch patterns are identical in two-syllable words, regardless of whether the first syllable has a schwa or some other vowel. This is confusing from an English or Dutch point of view, because a schwa is by definition unstressed and must therefore be unaccented. However, it makes perfect sense from the point of view of, say, Japanese, in which it is known that accent can occur on syllables that are voiceless or elided between voiceless obstruents (Sugito and Hirose 1988). If we simply abandon the assumption that the anchor point for tonal events must be a phonetically 'prominent' syllable, then the presence of a pitch accent on a schwa ceases to be a problem.

those found in music: for instance, dissociations between rhythmic prominence (defined by the underlying beat), and melodic or dynamic prominence (defined by harmonic changes, note durations, added loudness, etc.).

Certain syllables in a language like English or French are sometimes accompanied by pitch accents. However, pitch accents are elements of the intonation contour and do not in themselves represent the acoustic realisation of stress. They serve as an indirect cue to syllable prominence, because they must be associated with metrically strong syllables, but (contra the Bolinger–IPO view) they do not in and of themselves constitute the prominent syllable's prominence. In some languages (like English or Dutch), the metrically prominent syllables to which pitch accents are associated are also dynamically stressed. In other languages (like French or Indonesian) pitch accents may associate to syllables which are not necessarily stressed and which may not seem 'prominent' either to native speakers or to phonetically trained listeners.

2.3 Analysis of pitch accents in terms of level tones

So far we have discussed pitch accents as the most important of the phonological events into which pitch contours can be analysed. However, one of the most conspicuous features of AM descriptions of intonation is that in many cases pitch accents are themselves further analysed as sequences or combinations of High (H) and Low (L) tones. At first glance, this makes it appear that the AM theory takes a clear stand in what was long known as the 'levels-vs-configurations' controversy – the debate between those who analysed contours in terms of distinctive pitch levels and those who took the primitives of intonational analysis to be pitch movements or 'configurations'. In fact, I think it is fairer to say that the AM theory successfully resolved this debate.

2.3.1 The levels-vs-configurations debate

The levels-vs-configurations controversy was without a doubt the premier theoretical issue in intonation research from the early 1950s to the early 1980s. The issue was first posed in these terms by Bolinger (1951), who drew attention to a variety of problems with the then widely accepted American structuralist analyses of Pike (1945), Wells (1945), and Trager and Smith (1951). The most characteristic feature of these descriptions was that intonation was analysed in terms of four level 'pitch phonemes' (Low, Mid, High, and Overhigh),[6] which

6 The four-level analysis appeared in two versions, both of which used numbers to label the pitch levels. Pike's original version (Pike 1945, but presented in teaching materials for English as a foreign language as early as 1942) numbered the levels from top to bottom, so that Overhigh was 1, High was 2, and so on. The version first published in Wells 1945, which became the standard pre-generative American analysis when it was adopted by Trager and Smith (1951), numbered the levels from 1 at the bottom to 4 at the top. The essence of both analyses – in particular, the restricted distribution and special status of the highest level – is captured by the labels Low, Mid, High, and Overhigh given in the text.

were said to occur at certain structurally salient points in the utterance (see further Hockett 1955, 1958; Trager 1964). Bolinger pointed out that this four-level system is, in effect, both too rich and not rich enough: on the one hand, it predicts the existence of intonational contrasts that do not seem to occur; and on the other hand, it is unable to provide an analysis for various contours that clearly do occur. He attributed these problems to the use of pitch levels as primitives, and argued that the distinctive functional units of intonation were really 'configurations' like rise or fall.

In retrospect, it is clear that many of the problems Bolinger discussed were the consequence of certain general theoretical precepts of American structuralist linguistics, not just of the 'levels' idea (cf. Ladd 1983b). Throughout the debate, one of the problems that nagged at many 'configurationists' (e.g. Sledd 1955; Ladefoged 1967) was the need to be able to refer to actual F_0 levels in an explicit phonetic description of intonation, which seems to require some phonological notion of pitch level, and which suggests that there is some flaw in the dichotomy posed by Bolinger. Nevertheless, for roughly thirty years the levels-vs-configurations debate was an issue that intonation researchers felt they had to take sides on, and it certainly influenced the way individual researchers viewed their own work.

For example, the idea that intonation involves distinctive movements between relatively low pitch and relatively high pitch was developed independently during the 1960s by both Isačenko and Schädlich (1970) and the IPO researchers (see section 1.2). However, Isačenko and Schädlich's work emphasised the two levels; while the IPO work emphasised the movements between the two levels; and the IPO researchers seem not to have appreciated the extent of the common ground between their work and Isačenko and Schädlich's (see Cohen and 't Hart 1967: 189; 't Hart and Cohen 1973). While the isolation of Eastern European scholars during the Communist era undoubtedly played a role in this, it seems reasonable to suggest that the levels-vs-configurations debate was primarily responsible for Cohen and 't Hart's belief that their approach and Isačenko and Schädlich's were fundamentally opposed. In this connection I should note that the levels-vs-configurations issue was the basis of my own initial rejection of the autosegmental approach to tonal structure (Ladd 1978).

In any case, the AM theory effectively resolves this debate. Two theoretical claims are central to this resolution. These are, first, the reduction of the number of distinctive levels to two (viz. High and Low), and second, the acknowledgement of the existence of pitch accents.

First, by reducing the number of distinctive levels to two, the AM theory avoids the problem of predicting more contrasts than there are. For example, the four-level analysis of English predicts the existence of six distinct falling contours /21/, /31/, /41/, /32/, /42/, and /43/. It is almost certain that these sharp distinctions are of no relevance to the linguistic system of English, even if contours corresponding roughly to the six phonetic possibilities might be found. By analysing all falling contours as HL, and ascribing variations in the height of the H or the L to orthogonal factors, a two-level theory seems to strike the right balance between phonological contrast and paralinguistic variation. This point echoes one made by Bolinger himself (1951: 210, reprinted 1965: 16), though of course Bolinger was arguing *for* a configurations approach, not just *against* a four-level approach: '"231 and 241" (plus a note on synonymy) is less efficient than "rise–fall" (plus a note on pitch range), because the qualifying note in the latter can be generalized for all configurations and does not have to be repeated.'

Second, by acknowledging the existence of pitch accents *at some level of analysis*, the AM theory can use the primitive H and L tones to refer to pitch level in phonetic realisation, while at the same time acknowledging that pitch accents are in some sense distinctive pitch configurations. In this, the theory crucially diverges from the original 'levels' view, which directly analysed utterance contours in terms of levels.[7] In fact, with respect to pitch accents and phonetic realisation there is little practical difference between the AM view and the IPO theory. The IPO theory regards the pitch configurations as units, and emphasises the primacy of 'configurations'; but in modelling actual contours the theory nevertheless specifies starting and ending levels for each local configuration, defined with reference to the upper and lower declination lines. These starting and ending points are effectively equivalent to tonal targets in the AM approach.

The only important respect in which the levels-vs-configurations debate might still be said to be unresolved is in the realm of perception. One of the central arguments of the IPO researchers, for example, has always been that movements are crucial to the perception of intonational distinctions. We cannot hear that something is high, so the argument goes, except relative to something

7 This statement is not quite true of Pike, who assumes that the intonation of an utterance can contain a number of 'primary contours' consisting of two or more levels. These primary contours are somewhat comparable to AM pitch accents.

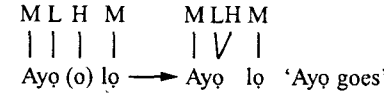
else in the context that is lower (cf. section 5.2.1). To some extent this is clearly true, but it seems likely that the case has been overstated. The best evidence of this is the existence of languages like Yoruba, in which it is uncontroversial that the system of lexical tones is based on distinctive levels: in an utterance that consists only of H tones, there is nothing lower in the context to serve as a point of comparison. A general (i.e. in this case non-Eurocentric) theory of pitch perception will need to make clear how listeners are able to perceive the 'highness' of the H tone in the absence of syntagmatic comparison to M or L. Once we have such a theory, it is plausible to assume that it will apply to English or Dutch as well. In any case, the argument from perception is not by itself enough to sustain the 'configurations' position in the levels-vs-configurations debate.

2.3.2 *The African example*

The development of ideas about intonational structure in the AM theory has clearly been influenced by the tonal phonology of African languages. In these languages it is well established that two 'level' lexical tones (such as H and L) can occur on the same syllable and yield a phonetically falling or rising contour. It will be useful to provide a few sample cases, so that readers unfamiliar with tone languages may have a clearer idea of the kinds of phenomena that are involved.

There are two main kinds of motivation for analysing syllable contours as sequences of level tones. The first type involves segmental elision. In many African languages, the vowel of a syllable may be elided phonetically while leaving its tone to be realised. In this case, one of the syllables adjacent to the elided syllable is realised with two tones. If the two tones are different, the result may be a contour or pitch change on the syllable. This can be seen in the following example from Yoruba:

(15)



Here, the H tone on *o* ('he/she/it') remains in the tonal string, even though the vowel to which it is underlyingly associated is completely elided.[8] The H tone reassociates to the preceding syllable *-yọ*, which then has two tones, L and H. This two-tone sequence is realised as a rise on *-yọ*.

8 The elision of the /o/ is probably a grammatical phenomenon rather than a synchronic phonological or phonetic process (see Rowlands 1969: 34f.), but that does not affect the point being made here.

The second motivation for analysing contours as sequences of levels in many African languages involves economy of description. For example, in Efik (as described by Ward 1933), monosyllabic verb stems have one of three distinct tonal patterns, High, Low, or Rise. Thus:

(15) dep (High) 'buy' du (Low) 'live' ka (Rise) 'go'

If we infer from this that there are three separate tone phonemes – High, Low, and Rise – then for two-syllable verb stems we predict nine possibilities: Low–High, Low–Rise, Rise–Low, and so on. In fact, there are only three possibilities for two-syllable stems, exactly as for one-syllable stems. These are High–High, Low–Low, and Low–High:

(17) kere (H–H) 'think' dori (L–L) 'put' fehe (L–H) 'run'

An obvious account of these facts is to say that there are only two tone phonemes, High and Low, and only three permitted tonal patterns for verb stems, namely High, Low, and Low–High. The surface realisation then depends on the association of the tonal pattern to the syllables available in the stem:

(18)

H	L	LH
dep	du	ka

H	L	L H
kere	dori	fehe

The description with only H and L tones has a smaller inventory of basic units, and predicts exactly which two-syllable patterns are possible.

Examples of these two kinds are to be found in language after language throughout sub-Saharan Africa (for good further background reading see Welmers 1973; Odden 1995; Heine and Nurse 2000), and formed a key impetus for the early development of autosegmental theory in works such as Leben (1973), Goldsmith (1976), and Williams (1976). This makes it clear that a place for 'level tones' must be found in our general understanding of linguistic pitch systems. Whether level tones are present in European languages is of course a separate question, to which we now turn.

2.3.3 *Phonetic evidence for distinctive levels*

While the example of African tone systems has undoubtedly influenced the development of AM intonational phonologies, there are other more direct

empirical foundations for the view that pitch movements may be represented as sequences of tones. The most important of these is Bruce's discovery that a precisely aligned peak, not a rise or a fall, is the most reliable correlate of word accent in Swedish (see sections 1.3.2 and 2.2.3). Certain pitch movements that might ordinarily be taken as distinctive – specifically, the rise on the accented syllable in citation forms with Accent 1 – are shown by Bruce's work to be *transitions* from one phonologically specified point to another. That is, the F_0 configurations that happen to span the accented syllables play no useful role in phonetic description of the overall contour; the invariant features of the pitch system appear to be the *turning points* in the contour rather than the transitions that connect them.

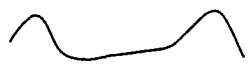
Bruce's data show very clearly that the 'word-accent maximum' – the highest F_0 value associated with the accented syllable – is the most reliable anchor for the phonetic description of contours: it always occurs at the same time relative to the accented syllable, and its F_0 level in a given utterance context is effectively constant. In order to account for these regularities, it is appropriate to treat the word-accent maximum as a phonological entity. In Bruce's words, we need to 'represent the tonal commands induced by the prosodic features in terms of two pitch levels, which are considered to be linguistically relevant: a LOW and a HIGH ... These tonal points can be viewed as ideal F_0-targets and give a rough picture of the essential tonal characteristics of an utterance' (1977: 130f.). Bruce concludes that in modelling accentual pitch configurations, 'reaching a certain pitch level at a particular point in time is the important thing, not the movement (rise or fall) itself' (*ibid.*: 132). He also goes on to state the principle that the actual phonetic scaling of the idealised Hs and Ls is a separate problem: 'this tonal representation has to be further elaborated to fit actual F_0 data ... by the operation of F_0 rules'.

For a long time, one of the most important arguments in favour of 'configurations' rather than 'levels' was that it is all a matter of how the analyst looks at the problem: in Bolinger's words (1986: 225f.), 'it makes no difference, in describing a movement, whether one says "first you are going to be up and then you are going to be down" or "you are going to go down" ' But the implication of Bruce's findings is that it does matter. Moreover, since Bruce's work, there is now abundant evidence that the pitch level of certain *target points* is highly significant communicatively. For example, Maeda (1976) found that the low pitch at the end of sentence-final fall varies little for any given speaker, and this finding has been replicated in many studies since (e.g. Menn and Boyce 1982; Liberman and Pierrehumbert 1984; Anderson and Cooper 1986; Connell and Ladd 1990; Ladd and Terken 1995; Shriberg *et al.* 1996 – the studies by Menn

and Boyce and by Anderson and Cooper are based on spontaneous or semi-spontaneous speech). Similar consistencies in the height of accentual peaks are less readily observed, because peaks can vary so much for other reasons (emphasis, place in utterance, etc.). In general, though, when other factors are carefully controlled, the height of peaks seems to be as regular as that of final low end points.

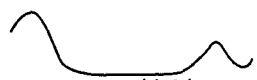
The study that drew attention to the regularity of peak height was Pierrehumbert's experiment on pitch range modification in two specific contours of English (Pierrehumbert 1980: ch. 3; Liberman and Pierrehumbert 1984: sect. 2). The two contours, and appropriate contexts for their use, are illustrated in the following exchanges:

(19) (a) '*Background–answer (BA) contour*'
Question: What about Anna? Who did she come with?



Answer: Anna came with Manny.

(b) '*Answer–background (AB) contour*'
Question: What about Manny? Who came with him?



Answer: Anna came with Manny.

As can be seen, both contours consist of two accents, but the accents differ according to whether the accented word is (in Pierrehumbert's terminology) the *answer* to the question or part of the discourse *background* that is presupposed. The tonal analysis of the two accent types is a point of considerable uncertainty (a point we return to in section 4.2.2), but what is important for the point under discussion here is that there is a consistent relation of pitch level between the peaks of the two accents: in the AB contour, the peak of the background accent on *Manny* is markedly lower than that of the answer accent on *Anna*, while in the BA contour the two peaks are approximately equal.

Pierrehumbert had speakers produce multiple repetitions of these contours, with ten different 'degrees of overall emphasis'. She then measured the F_0 levels of several presumed tonal targets, including the peaks of both accents. For each of the two contours, all four speakers showed virtually constant relationships

between the height of the two accent peaks, irrespective of the pitch range. This is shown in figure 2.2, in which data from both contours for one speaker are plotted. The plots for the other speakers are comparable. It is difficult – though not impossible – to avoid the conclusion that the relationship between the two peaks is being rather carefully controlled.[9]

While the *Anna/Manny* experiment focuses our attention on the *relation* between two peaks, the consistencies it reveals are evident even in studies where pitch range is not experimentally manipulated. For example, if we look at the mean values of sentence-initial accent peaks for individual speakers in the data reported by Ladd (1988), we find that standard deviations average only 7.5 Hz (ranging from 3.7 Hz to 11.5 Hz), suggesting rather considerable consistency in speakers' productions. To put this finding in perspective, it is instructive to compare it with results from a methodologically similar study of Yoruba (Connell and Ladd 1990), since we might expect that speakers of Yoruba, which has three lexical level tones, would control F_0 level rather carefully. Standard deviations on sentence-initial H-tone question particles in this study averaged 14.8 Hz, with a range from 8.3 to 31.4 Hz. We can conclude that, under roughly similar conditions, English speakers control pitch level at least as precisely as Yoruba speakers. Similar consistency in the final high pitch of questions in French is reported by Grundstrom (1973). All of this is what we might expect if, at some level of description, intonation in languages like English or French involves 'tones' that are comparable to those in the tone languages of Africa.

Nevertheless, the question is complex, because it is not necessarily easy to distinguish regularities of pitch level from regularities of pitch excursion (the size of a local pitch movement). That is, given a pair of contours of the idealised form shown in figure 2.3, it is equally plausible to describe the

9 In their model of these data, Liberman and Pierrehumbert attempt to show that in fact *both* contour types exhibit a single underlying relationship between the two peaks, despite the superficial difference. Specifically, they suggest that the surface patterns result from the combined effects of the answer–background ratio (which specifies that the 'answer' accent peak is higher than the 'background' accent peak by some fixed proportion) and 'final lowering' (which lowers the last accent in a sequence by a certain factor). That is, in the BA contour the second peak is underlyingly quite a bit higher than the first, but is lowered by final lowering to the point where the two are almost the same; in the AB contour, the second accent is already underlyingly lower than the first, and is lowered even more by final lowering. I think this is a mistaken interpretation (see Ladd 1990b; Calhoun 2006). However, the issue of whether we are dealing with one constant relationship or two does not affect the point being made here, namely that peak F_0 level is being carefully controlled for linguistic purposes.

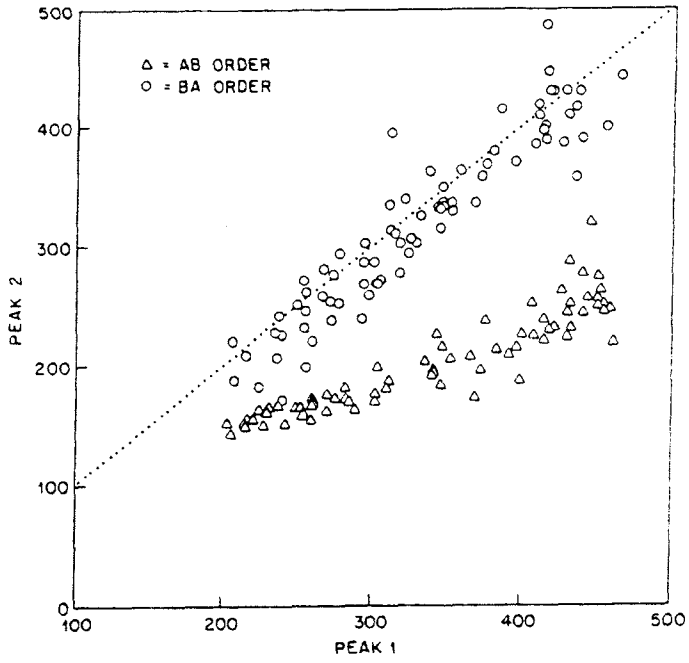


Figure 2.2 *Data for one speaker in Pierrehumbert's* Anna/Manny *experiment. For each individual utterance, the figure plots the F_0 value of the accent peak on* Anna *against that on* Manny. *The 'AB' order has the 'answer' before the background, while the 'BA' order is the reverse (see text for more detail). From Liberman and Pierrehumbert 1984.*

difference between a and b in terms of level – x′ is higher than x – or in terms of excursion size – (x′–y′) is greater than (x–y). In cases where the distinction can be made clearly, however, it appears that the level description better expresses the regularities in experimental data.

For example, perceptual experiments suggest that listeners attend to target level, and are, in fact, exceedingly poor at judging excursion size per se. An experiment by 't Hart (1981) is particularly interesting in this connection, as it was avowedly motivated by the theoretical assumption that excursion size is what is perceptually relevant. In 't Hart's words, 'insofar as pitch movements may lend prominence, one could ask, in the case of an utterance in which two syllables are made prominent by means of similar pitch movements, how much larger either of the two pitch movements should be in order to make that

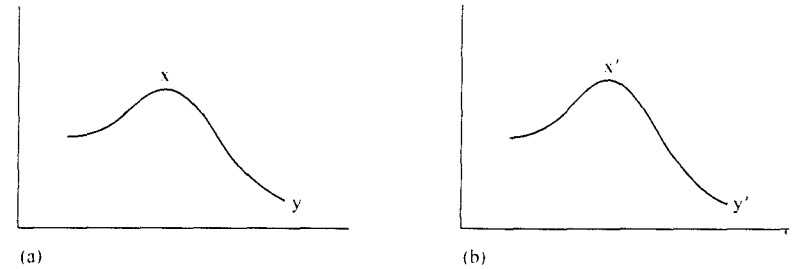


Figure 2.3 *Two different approaches to the quantitative definition of the prominence or pitch range of accent peaks. The relative prominence of the two idealised accent peaks in the figure could be described in terms of relative peak height (the difference between* x *and* x′*), or in terms of the size of the pitch excursion or amount of pitch movement (the difference between* x–y *and* x′–y′*). In many cases (as here) the two kinds of definitions are difficult to distinguish empirically.*

syllable more prominent than the other' (1981: 812). That is, 't Hart assumed that excursion size is correlated with degree of prominence, and set up his experiments to exclude the influence of peak height. Under these conditions, the subjects' performance was very poor. Other experiments in which listeners are able to compare peak heights directly (e.g. Rietveld and Gussenhoven 1985; Silverman 1987; Gussenhoven and Rietveld 1988; Terken 1991) result in quite fine discrimination performance. This suggests that relative height is not a distracting factor that needs to be controlled in designing experiments, as assumed by 't Hart, but is central to our ability to distinguish the relative prominence of accents. This idea has been extensively developed in the work of Laura Dilley (e.g. Dilley 2005; Dilley and Brown 2007).

Similar conclusions apply to the control of pitch in speech production, where we might wonder whether a speaker's intention is to produce a specific pitch level or a specific excursion size. For example, in Pierrehumbert's *Anna/Manny* experiment, as we just saw, the relationship of peak F_0 at the two accent peaks was constant under changes in overall pitch range, but Liberman and Pierrehumbert (1984: 210–15) also analysed the same data from the point of view of excursion size. That is, they looked at the relation between the excursion sizes of the two accent configurations to see if it was also constant. While there was, not surprisingly, a substantial correlation between overall pitch range and excursion size, the relation between excursion sizes in any given utterance was far less regular than the relation between peak heights.

2.3.4 Scaling of tonal targets as an empirical issue

If we accept the conclusion that distinctive pitch levels are involved in languages like English, we must acknowledge that we do not understand how the distinctiveness of pitch level works. One of the virtues of the AM approach is that it insistently draws our attention to pitch level as a problem for investigation.

Previously, one of the objections to analysing pitch contours in terms of levels, rather than configurations such as rise or fall, had been that the levels do not reflect phonetic reality. Bolinger (1951) suggested that the number of levels in any given analysis is merely a matter of how finely the analyst divides up what is essentially a gradient, while Lieberman (1965) showed that the pitch levels of the American structuralist analysis do not correspond systematically to actual observed levels in the instrumentally determined pitch contour. For the two-level analysis proposed by Bruce or Pierrehumbert, these objections are essentially irrelevant. With only two levels, the levels cannot possibly have any direct correspondence to phonetic reality; anyone can see that contours are more varied than that. Instead, H and L are phonological abstractions, comparable to phonemes, and there is no reason to expect them to be realised always in the same way. Rather, the phonetic realisation of H and L – like the phonetic realisation of any other phoneme – is subject to a variety of conditioning factors, which may make any given occurrence of H or L come out phonetically in a quite different way from some other occurrence. Consequently, the description as a whole must have theoretically interesting things to say about the mapping from abstract Hs and Ls to actual F_0 levels in contours. It thus seems fair to say that the phonological analysis of intonation in terms of only two distinctive levels implies a genuine attempt to come to grips with questions of phonetic realisation in pitch systems generally.

In any event, by describing contours in terms of only two distinctive levels, and by treating phonetic realisation – 'scaling' – as a separate problem, the AM view has virtually eliminated the levels-vs-configurations debate as an active issue in intonation research. In languages with lexically specified level tones, such as Yoruba, it is entirely uncontroversial that we need some sort of description with abstract phonemic levels and concrete phonetic realisation rules. The innovative idea in Bruce's and Pierrehumbert's work lay in seeing that the same idea of phonologically conditioned variation in the realisation of lexical tones can be extended without difficulty to the elements of a level-tone phonological representation of intonation. Once this is accepted, the basic opposition between high and low emerges naturally as a description of how

intonation works in many languages. In place of the unresolvable levels-vs-configurations conundrum, the AM theory has spawned the much more tractable empirical issue of how tonal targets are scaled.

In his dissertation, Bruce (1977: 131–43) makes some preliminary suggestions about the phonetic realisation of the H and L tones. Possibly to avoid straying too far from the received ideas of the Pike/Trager-Smith system, he suggests that the phonetics of intonation might be represented in an idealised way in terms of four F_0 levels, numbered from 1 (lowest) to 4 (highest). Any given H or L tone is realised at one of these levels. Bruce's discussion very explicitly brings out the distinction between abstract phonology and concrete phonetic realisation discussed in section 2.3:

> F_0-level 1 is considered to the base level and is the true representative of the LOW pitch level [i.e. L tone]. The F_0 movements can roughly be described as positive deviations ... from this base level ... In certain contexts the LOW pitch level will also be specified as F_0-level 2 (and occasionally as F_0-level 3). The HIGH pitch level [i.e. H tone] can be specified as F_0-levels 2, 3 or 4, depending on the context. This means that F_0-level 2 can represent both a HIGH and a LOW pitch level, which may seem paradoxical. But the pitch levels HIGH and LOW are to be conceived of as relative and contextually specified for each case as a particular F_0-level. (1977: 137)

Bruce goes on to note that he idealises the four F_0 levels as unchanging throughout the utterance, but he acknowledges that there is some evidence that they actually decline or drift lower.

In subsequent work (Bruce and Gårding 1978), the idea of four F_0 levels is dropped, and declination is built into the realisation model in terms of a distinction between 'statement lines' and 'focal lines'. The statement lines are a 'topline' and a 'baseline', which connect the peaks and valleys, respectively, of the *non-focal* word accents of the phrase; these lines decline gradually. The focal lines show the level of the word-accent maxima and minima for the *focused* word or words of the phrase; these lines are nearly horizontal, and roughly reflect the top and the bottom of the speaker's range for any given utterance. The Bruce and Gårding realisation model is shown in figure 2.4. By distinguishing the statement lines from the focal lines, the Bruce and Gårding model incorporates into its description of phonetic realisation a notion of what I have elsewhere (Ladd 1992) called *tonal space*. Tonal space is a subset of the overall speaking range which is available for realising tonal distinctions at any given point in the utterance. In Bruce and Gårding's model, the tonal space is defined by the 'statement lines', and changes throughout the course of the

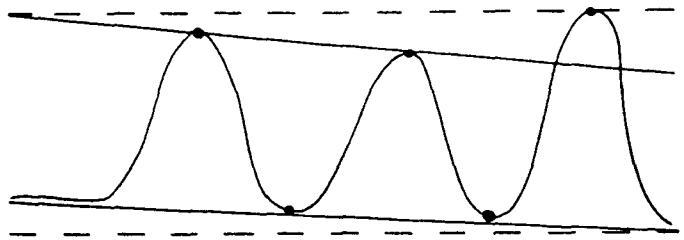


Figure 2.4 *Phonetic realisation of pitch features in the model proposed by Bruce and Gårding. The limits of the utterance range are shown by the horizontal 'focal lines'; and the range within which most pitch accents are realised is shown by the gradually declining 'statement lines'. After Bruce and Gårding 1978.*

utterance. The overall range is the one defined by the 'focal lines', which does not change.

The term 'tonal space' is my own, intended as a name for an informal construct that is central to the way AM intonational phonology treats phonetic data. 'Tonal space' is intended to echo 'vowel space', and has a similar – informal and essentially metaphorical – role in phonetic description: it reifies the limits on a set of observed values in phonetic data. When we say that different speakers have different vowel spaces, we mean that the observed acoustic details of corresponding vowels spoken by the two speakers differ in some consistent way. We also readily talk about the expansion or contraction of an individual speaker's vowel space as a function of speech rate, emphasis, and other factors. Informally, the correspondence between one vowel token and another is reflected in the fact that they occupy corresponding positions in two vowel spaces. Similarly, when we say that two different F_0 turning points both represent H tone, we may think of them both as being at the top of the tonal space for a given speaker at a given time in a given utterance; in effect, the phonetic realisation of pitch features is defined relative to the tonal space. Clements's 'tone level frame' (1979), Pierrehumbert and Beckman's 'transform space' (1988: 182), and 'register' (e.g. Poser 1984; Connell and Ladd 1990) are all instantiations of the tonal space idea. We return to discuss this idea in more detail in section 5.2.

An obvious difference between vowel space and tonal space, of course, is that tonal space is much more variable. It is well known that men, women, and children have different vowel spaces, but the differences are small and impressionistically inconspicuous; similarly, measurable differences between

corresponding vowels in stressed and unstressed syllables are difficult to detect impressionistically. So it is easy to take a phonetic symbol like [i] or [a] as having some sort of empirical validity across speakers and contexts. By contrast, the pitch differences between different voices, between different utterances, and between one part of an utterance and another are impressionistically obvious and often communicatively significant, so we are much more aware of the abstraction involved in identifying one pitch target with another. Nevertheless, if we allow that the actual phonetic value of the tonal space may change or evolve continuously during the course of the utterance (e.g. as a result of 'declination'), the tonal space idea may help us think about the relation between local tonal targets and overall shapes or trends in pitch contours. This is the topic of the next section.

2.4 Phonological interpretation of global F_0 trends

It has been clear at least since Pike (1945) that F_0 tends to decline over the course of phrases and utterances, both in tone languages and in languages like English or Dutch. It has also been noted that in a number of languages this 'declination' is suspended or reversed in questions (see e.g. the especially clear evidence from Hausa in Lindau 1986, and Inkelas and Leben 1990, as well as the data from Thorsen 1980 discussed in section 1.3.2). Observations like these are easy to incorporate into an overlay model of F_0, in ways that we reviewed in section 1.3.2. If we reject overlay models, we must nevertheless offer some account of such global trends.

The solution to this problem is in two parts. The first part involves drawing an explicit distinction between the abstract phonological specifications and their phonetic realisation. In the terms just introduced in section 2.3.4, that is, it involves treating the problem of declination and other global trends as part of the more general question of how tonal targets are scaled. The second part involves generating global trends by a sequence of local (i.e. phonologically controlled) modifications of the phonetic realisation parameters. Given the notion of tonal space, then, modifying phonetic realisation parameters in describing the phonetic detail of intonation may be thought of informally as making discrete localised changes to the tonal space. By modelling the continuous evolution of the tonal space during the course of an utterance by means of a sequence of localised changes, AM models avoid the problem of quantitatively describing overall contour shapes (e.g. Gårding's 'grids').

The difference can be seen clearly in the treatment of declination. In many past quantitative descriptions of declination (such as the IPO model,

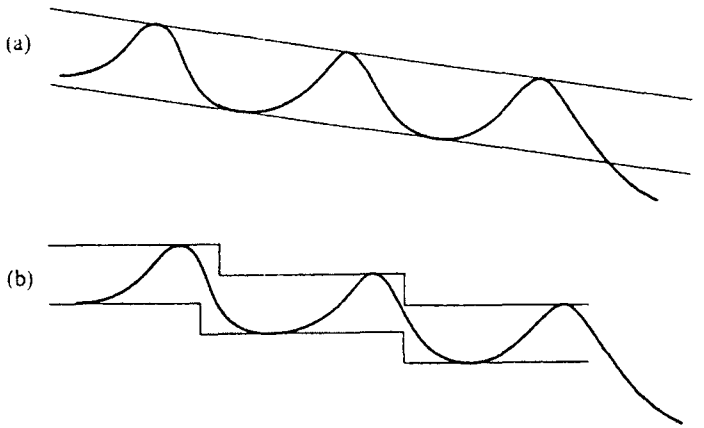


Figure 2.5 *Two different ways of describing the overall downward trend of a pitch contour. Idealised contour and 'tonal space' diagrams show the overall downward trend as the consequence of (a) gradual declination or (b) downstep at specific points in the utterance.*

or Fujisaki's model), the downward slope of the contour is directly specified as such. In Fujisaki's model, the phrase component is a continuous function generated separately from the local accentual events; in the IPO model there is a standard formula for computing the starting level and the rate of declination, based on the length of the utterance ('t Hart 1979). In all AM work since Pierrehumbert 1980, by contrast, the evolution of the tonal space is treated in a way more in keeping with the idea of a sequence of localised phonological events. Specifically with regard to declination, Pierrehumbert (1980) advanced the hypothesis that much of declination can be accounted for as the result of *downstep*[10] – the stepwise lowering of pitch (or of the tonal space) at specific pitch accents. Figure 2.5 shows the difference between describing an idealised downward trending contour in terms of global declination and treating it as the result of the repeated localised occurrence of downstep.

This shows how an apparent global trend in a pitch contour can be analysed in a way more in keeping with the assumptions of AM intonational phonology.

10 Beckman and Pierrehumbert (1986), following Poser (1984), replace Pierrehumbert's term *downstep* with the coinage *catathesis*, but in Beckman and Pierrehumbert 1992 they return to *downstep*, and this usage now appears to be well established. The neologism was originally motivated by a desire to steer clear of disagreements within the Africanist literature over distinctions like *downstep* vs *downdrift* and *automatic* vs *non-automatic downstep*.

However, it should be emphasised that the notion of downstep is not simply a clever trick motivated by the search for a way to avoid having an 'overlay' component in the phonetic realisation model. Rather, inspiration for the notion of downstep in English comes from the tonal systems of many languages of sub-Saharan Africa. In many such languages, the second High tone in a sequence High–Low–High is realised at a lower level than the first High, and the level of the second High sets a new ceiling for the realisation of High tones until the end of a phrase or some other relevant prosodic unit. Similar phenomena have since been shown to occur in Japanese (Poser 1984; Kubozono 1989), conditioned by the presence of lexical accent: the H tone of a phonological word, whether accented or unaccented, will be lower if it follows an accented word than if it follows an unaccented word. The use of local downstep to generate global trends in English entails the claim that English has a distinction between downstepped and non-downstepped accents: a downstepped accent is realised at a lower F_0 level than a corresponding non-downstepped one.

Justification for drawing the distinction between downstepped and non-downstepped accents in English can be demonstrated most clearly on the basis of short utterances with only two accents. Consider a phrase like *my mother's diaries*. First, recall (section 1.1.2) that there is a distinction between the two possible prominence patterns strong–weak and weak–strong (i.e. *my MOTHER'S diaries* and *my mother's DIARIES*). In the strong–weak version, the focus is on *mother's* (e.g. my mother's diaries, not my father's), while in the weak–strong version, the focus is either on *diaries* (my mother's diaries, not her scrapbooks) or on the whole phrase (my mother's diaries, not any of the family furniture). Now, with an ordinary declarative intonation, the pitch contour on the strong–weak version will show a peak on *mother's*, and *diaries* will be realised quite low in the speaking range:

(20)

my mother's diaries

But there are two distinct ways of realising the weak–strong version that both nevertheless remain roughly within the realm of ordinary declarative intonation. First, we can have

(21)

my mother's diaries

with the accentual peak on *diaries* equal to or higher than that on *mother's*. But second, we can also have

(22)

my mother's diaries

which has a downstepped accent on *diaries*. Though (22) is very similar phonetically to (20), there is an unmistakable difference: (20) focuses on *mother's*, and (22) does not. Phonetically, that is, a downstepped accent on *diaries* is very similar to the absence of accent; but pragmatically it is clearly an accent.[11]

In order to account for these observations, we need to be able to say that *diaries* in (22) is metrically stronger than *mother's*, even though it is scaled considerably lower. This is achieved by postulating a distinction between downstepped and non-downstepped accents. The downstepping adds a nuance like finality or completeness, but does not make the accent 'less prominent' in the way it affects the focus of the phrase: both downstepped and non-downstepped accents can manifest the underlying prominence pattern weak–strong. During the 1980s there was considerable debate over the appropriate phonological analysis of English downstep, but Pierrehumbert's basic descriptive insight – that there is such a distinction – has been widely accepted (see e.g. Hirst 1983; Ladd 1983a; van den Berg, Gussenhoven, and Rietveld 1992; a dissenting view is presented in Dainora 2001, but Yoon 2007: sect. 6.3 supplies a rejoinder). This, in turn, makes it plausible to account for at least some global downtrends in English in terms of the *repeated* occurrence of downstep.

Pierrehumbert and her colleagues (Pierrehumbert 1980; Liberman and Pierrehumbert 1984; Pierrehumbert and Beckman 1988) have devoted a great deal of attention to the detailed phonetics of downstep in English. The details of the model have changed – in particular, the 1988 version of the model much more clearly incorporates the tonal space idea than the original 1980 model – but one of the main points remains unchanged: namely that, given an appropriate scale in which to express F_0 values, the steps in a downstep sequence like that shown in figure 2.5 are of equal size. More precisely, the

11 Nevertheless, the phonetic similarity may have some effect on the pragmatic interpretation. Stefan Baumann (2006) shows clearly that, at least in German, downstepped accents are in some sense pragmatically intermediate between non-downstepped H* accents and absence of accent. Specifically, downstepped accents are often used on expressions that refer to entities that are inferable or otherwise not completely new to the discourse, whereas non-downstepped accents are more often used with expressions referring to completely new entities, and deaccenting tends to be reserved for entities that have been mentioned in the immediately preceding context.

value of each accent peak in a downstep series is a *constant proportion* (in terms of the model's parameters) of the previous peak. Similar results have since been obtained for other languages (e.g. Dutch: van den Berg, Gussenhoven, and Rietveld, 1992). Related work by Truckenbrodt (2002) has also shown that the size of the downsteps within a phrase in German predicts the size of the *upsteps* at the end of the phrase when it is non-final in the utterance.

These are important findings, because, like Bruce's data on Swedish word accents, or like the results of the *Anna/Manny* experiment, they are consistent with the view that target levels are actively controlled by the speaker and should therefore be represented somehow in a phonological description of intonation. These findings are also important because they call into question a significant motivation for overlay models: if target values can be generated from left to right, with only a small 'window' looking back to a previous value, it is possible to generate overall trends in contours without incorporating them as actual components of the model. Truckenbrodt's findings are especially revealing in this way, as they suggest that the phonetic realisation parameters for any given accentual pitch target can be specified locally on the basis of the position of the accent in a hierarchical prosodic structure, and because they can predict the size of both upsteps and downsteps – not merely an overall declination slope – using the same basic information.

In fact, the accent-by-accent model is discriminating enough to permit Liberman and Pierrehumbert to distinguish downstep from what they call 'final lowering' in production data. In an experiment in which speakers produced utterances containing a series of from two to five downstepping accents, Liberman and Pierrehumbert found that the final accent was invariably lower than would have been predicted by the constant proportion in the phonetic model of downstep. The amount by which it was lower was consistent across utterances and across experimental conditions. Liberman and Pierrehumbert therefore suggested that the scaling of the final accent was subject to two distinct effects, downstep and final lowering, each of which makes its own distinct contribution. This interpretation of the instrumental data is in line with observations of tone languages, in which the last tone in an utterance is often realised at a much lower level than would otherwise be expected (cf. Pike 1948: 28, and, for much more detailed acoustic data, Herman 1996). It is difficult to see how such specific effects on specific accents could be accounted for in an overlay model with anything like the same simplicity.

To be sure, Liberman and Pierrehumbert's specific conclusions have been called into question by Grabe (1998: ch. 6), who argues that the greater lowering on the final accent is due to the fact that the last two accents in Liberman

and Pierrehumbert's experimental materials are separated by a greater time interval. Her own results seem to suggest that the amount of downstep from one accent to the next is correlated with the temporal distance between the two accents, which would make sense if we were dealing with a global time-dependent downtrend rather than the accent-by-accent modification of one or more phonetic realisation parameters.[12] However, her conclusions may be affected by the fact that her experiment was done in British English, and Liberman and Pierrehumbert's in American English; a more recent study by Arvaniti and Godjevac (2003), based on both American English and Greek, seems to be broadly in line with Liberman and Pierrehumbert's idea of two distinct lowering factors at specific points.

It can hardly be suggested that the best way to model overall trends has been settled, or that declination is now fully understood. Nevertheless, as with some of the other issues discussed in this chapter, the AM perspective has made it possible to move beyond irresolvable theoretical standoffs toward tractable empirical questions about how tonal targets are scaled. At least to that extent, the AM view has led to genuine progress; some details of this progress are presented in section 5.2.

2.5 AM description and intonational universals

Now that we have laid out the four basic tenets of the AM theory and provided some justification for each of them, it will be useful to conclude with a concrete illustration of the theory's general descriptive potential. By discussing a specific example in some detail, I will show how the AM approach makes it possible to undertake meaningful cross-language comparison of intonational features. This is important, because cross-language comparison is prerequisite to understanding intonational universals.

There are enough widely shared properties of intonation systems that many investigators have hypothesised a universal common core of intonational signalling, related either to the control of respiration during speech production (e.g. Lieberman 1967) or to supposedly biologically determined communicative effects of higher and lower pitch (e.g. Bolinger 1978, 1986: ch. 9, 1989; Ohala 1984, 1994; Gussenhoven 2004: ch. 5). The tendencies generally emphasised by those who assume a universal common core for intonation

12 Actually, there is nothing in principle to rule out the existence of both downstep and 'true' (i.e. time-dependent) declination, and Pierrehumbert and Beckman (1988: ch. 3) incorporate this possibility into their model of Japanese. However, the empirical difficulties of testing such a model – distinguishing downstep from declination, and distinguishing both from natural variability in F_0 – are considerable, and the issue is by no means settled.

include declination (and more generally the association of low or falling pitch with completion); the association of high or rising pitch with both questions and non-finality; and the presence of local pitch movements on new or otherwise informative words. While it would be foolish to deny the existence of broad generalisations about how intonation affects the meaning of utterances, there are good reasons for scepticism about treating them as the basis for comparative description of intonation, that is, as a backdrop against which language-specific differences can be seen as minor variations on a single theme.

There are several problems with using universal tendencies as the basis for describing the intonation of specific languages. First, there are many obvious differences between languages, and between different dialects of the same language, which are often very striking to speakers of languages or dialects in regular contact. These are difficult to accommodate in any universal-based descriptive framework. Second and more important, many 'universal' generalisations are arguably based on a fairly Eurocentric sample, and are thus not really about linguistic universals but only linguistic areas. Particularly interesting in this connection is the work of Annie Rialland on question intonation in Africa, which provides clear evidence of a large group of languages in which the end of question utterances tends to be characterised by some or all of the following features: breathiness, slowed tempo, open vowels, and low pitch (Rialland 2007). Finally and most important, many universal generalisations are so broad, or so vague ('high or rising pitch'), that it is virtually impossible to subject them to any sort of meaningful empirical test. Without a generally agreed framework for describing intonation, it is difficult to compare intonation across languages and come to any reliable conclusion about the significance of the similarities, or of the differences. With such a framework, however, meaningful comparison ought to be possible.

As an example of the kind of comparison that is made possible in an AM analysis, I will here review the facts of Hungarian question intonation,[13] based on the presentation in Ladd 1983a. These facts have been fairly well established for several decades (for example, they were very clearly laid out by de Sivers 1965), but they have often been presented as extremely puzzling (e.g. by Gósy and Terken 1994) because of the conspicuous differences between

13 Note to readers who may be disturbed by the term 'question intonation': I am well aware that there is no simple correspondence – much less a deterministic relationship – between sentence type and tune (see e.g. Scherer, Ladd, and Silverman 1984; Haan 2001; Lickley, Schepman, and Ladd 2005). The term is intended only to refer to the fact that, in Hungarian, as in many other languages (including Greek and Italian), there is no usual lexical or syntactic indication of whether an utterance is a statement or a question, and in many cases intonation alone is used to signal the distinction.

the Hungarian pattern and question intonation in many Western European languages. It is not overstating the case to say that many Hungarian questions sound like emphatic statements to native speakers of English, which is prima facie something of a problem for a straightforward theory of intonational universals. Less conspicuously, the relation among lexical stress, focus, and pitch excursion is quite unlike what one might be led to expect on the basis of the Western European languages alone.

From an AM viewpoint, there are two major differences between Hungarian question intonation and the kinds of rising contour one finds in English or French. First, in Hungarian (as in many languages, including Russian, Romanian, and Greek) the neutral location for the nuclear accent in a yes–no question is on the finite verb.[14] Second, though the pitch on the verb is low, the question tune ends with a high-falling pitch movement, which sounds to speakers of many Western European languages like a declarative falling accent, but which is not necessarily even associated with a stressed syllable. Dealing with the second of these differences first, we may say that the Hungarian question tune consists minimally of an L pitch accent followed by an H_L edge tone sequence. The L* is associated with the lexically stressed syllable (which is invariably the initial syllable) of the nuclear-accented word; if there are at least two syllables following the nucleus, the H edge tone is associated with the penultimate syllable, and the L edge tone with the final syllable. For example:[15]

(23) L* H L
Beszél a tanár?
'Is the teacher talking?' (lit. talks the teacher)

In this example it happens that the penultimate syllable is also the stressed (initial) syllable of *tanár*, but that is coincidence, as we can see if we substitute a three-syllable noun:

(24) L* H L
Beszél a miniszter?
'Is the minister talking?'

14 The use of the terms 'nucleus (of the intonation contour)' and 'nuclear accent' for the most prominent accent of the phrase – called elsewhere 'focal accent' or 'sentence accent' or simply 'primary stress' – is discussed in section 4.1.1. The notion that 'neutral' accent signals that there is 'no special focus' is discussed extensively in chapter 6, as is the fact that the nucleus is located on the verb in yes–no questions in some languages.

15 The acute accent in Hungarian orthography indicates vowel length, not stress; lexical stress is fixed on the first syllable of the word. The L* indicates the low nuclear pitch accent. Other details of AM notation will be introduced in the next chapter.

If the penultimate syllable is also the nuclear syllable (e.g. if the verb is final, or if the question intonation applies only to a noun phrase), the whole tune is 'squeezed' to the right, so that the penultimate syllable is low-rising and the final syllable falling (what happens on monosyllabic questions will be discussed in 5.1.3). We might represent this as follows:

(25) L* HL
A tanár?

If there are prenuclear accents, they are normally medium-range H peaks:

(26) H L*H L
A magyar nehéz nyelv?
'Is Hungarian a difficult language?' (lit. the Hungarian difficult language)

Now let us consider the first difference mentioned above, namely that the 'neutral' location for the nucleus in a question is on the verb. Differences in focus and emphasis can be conveyed both by changing word order (as in Italian or Russian) and by changing the location of the nucleus (as in English or Dutch), but the nucleus goes on the verb if no special focus is intended. Thus:

(27) (a) L* H L
Vettél szódát?
'Did you buy soda (= mineral water)?' (neutral, out of the blue)
(b) L* H L
Szódát vettél?
'Did you buy SODA?' (i.e. 'Was it soda that you bought?')
(c) H L*HL
Szódát vettél?
'Did you buy SODA?' (i.e. 'What about soda? Did you buy that?')

The interaction of nucleus shift and word-order change in these examples to convey different pragmatic messages is typical of Hungarian. The details are beyond the scope of this discussion, but it is worth pointing out that the fourth logical possibility, namely *vettél szódát* with the L* question nucleus on *szódát*, strikes native speakers as odd. The problem is that the intonation signals narrow focus (see section 6.1) on *szódát*, but the syntax requires narrow focus constituents to occur before the verb; there is thus a mismatch between the syntax and the intonation. For more on this general topic see, for example, Kiefer 1967; Horvath 1986; Vogel and Kenesei 1987; Varga 2002: ch. 6; Szendrői 2003; Wedgwood 2005.

It is not difficult to imagine how a phonetician of the 1960s, equipped only with a basic non-autosegmental idea of 'stress' as pitch prominence and of

'intonation' as global utterance contour, might be mystified by the effects just summarised. The contour in (27a), in which the phonetically most salient pitch excursion occurs on *szó-*, is precisely the contour that does *not* focus on 'soda': if we want to focus on 'soda' we have to use a different contour, such as the ones in (27b) and (27c). However, once we understand that we are dealing with a tonal sequence L*_H_L that does not occur in most Western European languages, we can see that all the facts about focus and the location of the phrase-final pitch peak are exactly what we would expect. The basic structural division of the contour into accent tones and edge tones, the way in which the tune is associated with major stressed syllables and boundaries, and the consequent variation in the realisations of the tune depending on the number and location of major stressed syllables and the location of focus, are just like what we find everywhere else. The only differences between Hungarian question intonation and other better-known intonation patterns in the European languages are that (a) it involves a different tonal sequence and (b) 'neutral' accent is on the verb.

On the other hand, it is worth emphasising that both the neutral accent placement, and the question tune itself, really are different from what we find in most Western European languages. That is, while the AM analysis brings out the basic features that the Hungarian question tune shares with many others, it also makes it possible, in a principled way, to avoid treating all question contours as variations on the same theme. If we simply say that questions universally have high or rising pitch at or near the end of the utterance, then Hungarian question intonation could be (and has been; cf. Bolinger 1989: 57ff.) treated as evidence for the universal nature of question intonation. In the light of the foregoing discussion, this seems unsatisfactory, and the AM analysis makes it possible to combine an explicit account of the phonology and phonetics of Hungarian question intonation with falsifiable statements about what is and is not an intonational universal. Ideas like Bolinger's or Ohala's are not ruled out by the AM approach – the work of Carlos Gussenhoven and his students shows that it is possible to combine theorising about the universal 'biological codes' in human speech with careful and detailed investigation of the intonation of individual languages and language varieties (e.g. Chen, Gussenhoven, and Rietveld 2004) – but they are constrained to rest on a firmer empirical footing that acknowledges differences where they occur.

II Pitch

3 *Analysis and transcription of intonation*

Now that we have explored the theoretical and empirical foundations of the AM approach to intonation, and illustrated its potential for meaningful comparison across languages, we can move on to consider technical details of the AM description of the intonational phonology of specific languages. The obvious starting point for any such discussion is Pierrehumbert's analysis of English. There are actually three distinct versions of this analysis: the original version presented in Pierrehumbert 1980; a revised version, intended to supersede the original, developed by Pierrehumbert in collaboration with Mary Beckman (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988); and a modified and simplified version of the revision that forms part of the original ToBI transcription system for Standard English (Silverman *et al.* 1992; Pitrelli, Beckman, and Hirschberg 1994; Beckman and Ayers Elam 1993; Brugos, Shattuck-Hufnagel, and Veilleux 2006). These are presented in section 3.1. The presentation is fairly summary; at several points brief reference is made to matters of disagreement, unresolved issues, other languages, and further developments, and many of these issues are taken up again in greater depth later on. In section 3.2 we consider the extension of AM principles to the description of languages other than English.

3.1 The Pierrehumbert analysis of English intonational phonology

3.1.1 Notational conventions

It will be useful to begin with a few remarks about notation, as it is important to distinguish mere notational conventions from the theoretical innovations that the notation system expresses. In keeping with the ideas presented in section 2.1, Pierrehumbert's notation represents the contour as a string of pitch accents and edge tones. All pitch accents consist of a single H or L tone, or a combination of two tones. In bitonal accents, one of the two tones is assumed to be central in some way; the central tone of a pitch accent is indicated with an asterisk, as either H* or L*, and is therefore generally referred to as the 'starred tone'.

In addition to this starred tone, a pitch accent may (but need not) contain a 'leading' (preceding) or 'trailing' (following) tone – for example, L+H* (High starred tone with a leading Low). The bitonal notation is used, for example, in cases where the pitch accent is characterised by rapid local F_0 movement rather than just a local maximum or minimum.[1]

Edge tones are divided into two types in Pierrehumbert's original analysis: *phrase accents* (notated H- and L-, or simply H and L) and *boundary tones* (notated H% and L%). Boundary tones in this strict sense are single tones – either High or Low – associated with the very end (or, in some cases in some languages, the very beginning) of an intonational phrase. Phrase accents (sometimes also referred to as 'phrase tones') are free-standing unstarred tones (i.e. unstarred tones that do not serve as leading or trailing tone in a bitonal pitch accent) that occur after the last pitch accent. In Pierrehumbert's original analysis, every intonation phrase ends with a sequence of a pitch accent, a phrase accent, and a boundary tone. The revised version of the analysis posits an additional layer of structure, the *intermediate phrase*, such that each intonation phrase consists of one or more intermediate phrases (for discussion see Beckman and Pierrehumbert 1986: sec 4.3). Given this structural distinction, the boundary tone (T%) is seen as the edge tone for the intonation phrase (*IP*), while the phrase accent (T) is said to be the edge tone for the intermediate phrase (*ip*). In an *IP* consisting of a single *ip*, there is no practical difference between the revised analysis and the original – the contour still ends with a sequence of a pitch accent, a phrase accent, and a boundary tone – but in an *IP* consisting of more than one *ip*, the non-final *ip*s end with a sequence of pitch accent and phrase accent only. There has been extensive investigation of phrase accents since the mid 1990s (see in particular Grice, Ladd and Arvaniti 2000; and sect. 4.1.4).

3.1.2 *Structure of tunes and basic intonational taxonomy*

The notation just summarised expresses several of the theoretical claims about intonational phonology that were outlined in chapter 2. First, and most obviously, it embodies the basic phonological claim that contours are to be analysed

1 In the original notation, leading and trailing tones (and also phrase accents, discussed immediately below) are written with a following raised hyphen (H⁻ or L⁻). The two tones of a bitonal pitch accent are joined with a plus sign (e.g. L*+H⁻). In some early work based on Pierrehumbert (e.g. Gussenhoven 1984; Féry 1993; Grice 1995a), both the plus sign and the hyphen are dispensed with, and one writes simply L*H. Others (e.g. most ToBI systems) retain the plus sign but omit the hyphen in bitonal accents, while retaining the hyphen for indicating phrase accents. In any case, for typographical simplicity the raised hyphen has tended to become an ordinary hyphen if it is used at all.

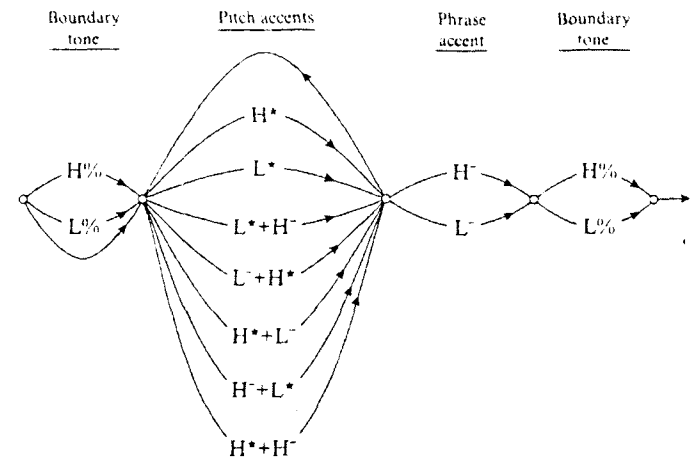


Figure 3.1. *Finite-state grammar to generate tunes of English intonation according to Pierrehumbert's original analysis. After Pierrehumbert 1980.*

as strings of elements occurring at well-defined points in the utterance. Second, and more specifically, it clearly distinguishes between two types of elements, pitch accents and edge tones. Third, it further analyses pitch accents as consisting of one or more H and L tones.

The notation also embodies several ideas about the *structure* of tunes. These are summarised by Pierrehumbert (1980) in the finite-state grammar shown in figure 3.1, which can be used to generate all the legal tunes of English. This grammar says that tunes are made up of one or more pitch accents, followed by an obligatory phrase accent and an obligatory boundary tone. This implies two interrelated theoretical claims about the structure of tunes that are worth highlighting here. First, the grammar implies that all possible sequences of pitch accents and edge tones are legal, an implication that Pierrehumbert explicitly endorses (1980: 30, published 1988: 13).[2] Second, it implies that there is no

2 Partly because of the theoretical claim that sequences have no special status, there is no standard notation for indicating sequences or combinations of tonal elements. In what follows I have adopted the convention of using underscores to separate the elements of sequences involving edge tones (e.g. H_L%, H*_L, L+H*_L_H%). The main motivation for this is that these sequences may be seen, at some level of analysis or in some descriptive framework, as combinations or units. (Most obviously, the combination of pitch accent and phrase accent, and in some cases boundary tone as well, corresponds to the nuclear tone of the British tradition.) In the few cases where I have had occasion to refer to sequences that would not normally be regarded as higher-level units or combinations (e.g. sequences of pitch accents, or sequences of tones in lexical tone languages), I have separated the elements with ellipses (e.g. H* ... H* for a sequence of high pitch accents).

internal constituent structure to the contour, in particular no analogue to the 'head' and 'nucleus' of the traditional British analysis. Together, these mean that there is no difference between 'prenuclear' and 'nuclear' accents, except – trivially – their position: for Pierrehumbert, the 'nuclear accent' is merely the last accent of the phrase. However, if we accept Beckman and Pierrehumbert's revised understanding of the phrase accent, then 'nuclear accent' can be defined as the last accent of the *intermediate* phrase. This usage has become increasingly common, and is made explicit in the ToBI training materials (Beckman and Ayers Elam 1993: sect. 1.4; Brugos, Shattuck-Hufnagel, and Veilleux 2006: sect. 2.3), which tell the user to think of the nuclear accent in this sense as more prominent than prenuclear accents. We return to this issue in section 4.1.1.

Despite the obvious innovations, there is much in the basic Pierrehumbert taxonomy of English intonation that is comparable with earlier analyses, in particular those of the British tradition. While denying any theoretical status to the notion of nuclear accents, Pierrehumbert does implicitly acknowledge that combinations of the final pitch accent and following edge tones correspond closely to traditional British 'nuclear tones'. Specifically, in the Appendix to the Figures of Pierrehumbert 1980, she illustrates the realisation of the twenty-two[3] occurring combinations of pitch accent, phrase accent, and boundary tone in a way that invites comparison to the categories of the British tradition. Table 3.1 summarises the twenty-two combinations and gives a possible British-style description of each. Some of the equivalences suggested are open to argument, but the existence of clear correspondences in most cases can be taken as some indication of the validity of the basic taxonomy of intonational contrasts. As noted earlier (section 2.1), this point was also made by Roach (1994) in his attempt to provide automatic conversion from traditional British 'tonetic' notation to ToBI notation.

Nevertheless, it is pointless to attempt to state a complete correspondence between the two systems. For one thing, the British tradition is not monolithic, but includes a range of divergent analyses of certain phenomena, including the general problem of 'heads' (patterns of prenuclear accents) and specific distinctions such as that between 'high fall' and 'low fall' (see the discussion in section 1.4). This means that there is no single agreed British inventory of

3 The original Pierrehumbert analysis posited seven pitch accent types, and there are actually twenty-eight logically possible combinations of seven pitch accents, two phrase accents, and two boundary tones. However, six of these are said to be indistinguishable from other sequences as a result of neutralisation brought about by the implementation rules. Four of the six involve the controversial H*–H accent discussed in section 5.2.5.

Table 3.1. *Correspondences between Pierrehumbert 1980 and British-style nuclear tones. In the Beckman and Pierrehumbert 1986 version of the analysis, the two contours marked with an asterisk would be variants of the calling contour, because of the modification of the downstep rule (cf. section 3.1.3). For the same reason, the contour marked with a dagger would be a high-range fall–rise in the 1986 version*

Pierrehumbert	British-style
H* L L%	fall
H* L H%	fall–rise
H* H L%	stylised high rise
H* H H%	high rise
L* L L%	low fall
L* L H%	low rise (narrow pitch range)
L* H L%	stylised low rise
L* H H%	low rise
L+H* L L%	rise–fall
L+H* L H%	rise–fall–rise
L+H* H L%	stylised high rise (with low head)*
L+H* H H%	high rise (with low head)
L*+H L L%	rise–fall ('scooped')
L*+H L H%	rise–fall–rise ('scooped')
L*+H H L%	stylised low rise*
L*+H H H%	low rise†
H+L* L L%	low fall (with high head)
H+L* L H%	low rise (with high head)
H+L* H L%	stylised low rise (with high head)
H+L* H H%	low rise (high range)
H*+L H L%	stylised fall ('calling contour')
H*+L H H%	fall–rise (high range)

nuclear tone types that we can compare item by item against the Pierrehumbert analysis. More importantly, the AM approach is intended to provide a new foundation for analysing intonational distinctions, not merely a new notation for the same old description. The divergence between the two approaches is most obviously reflected in the way the nuclear tone types are grouped in Table 3.1: these groupings make sense in the Pierrehumbert system, but not in the British system. The grouping based on the Pierrehumbert analysis shows five completely parallel sets of four types, plus two additional ones, whereas from the point of view of the British tradition, certain types like 'low rise' and 'high rise' show up rather unpredictably at several different places in the table, and

references to pitch range or to the type of preceding 'head' are required here and there to describe certain distinctions.

3.1.3 Accent types

Pierrehumbert's original analysis of English intonation posited seven possible pitch accent types: H*, L*, L+H*, L*+H, H+L*, H*+L, and H*+H. In the revised standard (Beckman and Pierrehumbert 1986) the H*+H accent was eliminated (see section 5.2.5), and the presentation in this section is based on the remaining six types. Of the six, two are monotonal (H* and L*) and four bitonal (L*+H, L+H*, H*+L, and H+L*). Three of the six have a fairly straightforward phonetic description – H* is a local peak, L* a local valley, and L*+H a rise from a low accented syllable – while the other three (L+H*, H*+L, and H+L*) require some comment, to which we turn below. However, even for the first three, readers who are familiar with the British or IPO traditions of transcription should bear in mind that the phonetic descriptions refer only to the accented syllable, and that there is no distinction in the theory between the 'nuclear' and the 'prenuclear' uses of these accents. Combining these accents with different edge tones can give rise to impressionistically very different contours. For example, when H* is followed by an L phrase accent, the result is a 'falling nuclear tone' (Halliday Tone 1) or an IPO 'Type A Fall', whereas when they are followed by an H phrase accent, the result is a 'rising nuclear tone' (Halliday Tone 2) or an IPO 'Type 1 Rise'. When they are not followed by any phrase accent, they mark prominent prenuclear syllables (e.g. Halliday's salient but non-tonic syllables).

Figure 3.2 shows nuclear H* accent followed by L phrase accent (panel a) and H phrase accent (panel b); both panels also show the H* in prenuclear position.[4] Figure 3.2 also illustrates a peculiarity of the H* accent in Pierrehumbert's analysis, which is that two H* accents in sequence are not connected by a straight transition from one peak to the next; instead, the transition is said to 'sag', and the second H* accent is realised as a local jump to the target pitch peak (Anderson, Pierrehumbert, and Liberman 1984). We will return to the local jump in discussing the L+H* accent below, and will revisit the sagging transition in section 4.1. Figure 3.3 shows a prenuclear L*.

The difference between H* and L*+H is clearest when they are followed by an L phrase accent, as shown in figure 3.4. This distinction appears to correspond to the distinction between what some versions of the British

4 The figures in this section are provided primarily to illustrate key points of the exposition, and are not intended as a complete introduction to Pierrehumbert's treatment of the full range of English contour types. Such an introduction is in any case more effectively presented in conjunction with sound files and interactive displays of acoustic analysis, and the reader who wants more examples and more detailed explanation is referred to the ToBI website (Veilleux, Brugos, and Shattuck-Hufnagel 2007), especially the detailed labelling guidelines (Beckman and Ayers Elam 1993) and training materials (Brugos, Shattuck-Hufnagel, and Veilleux 2006).

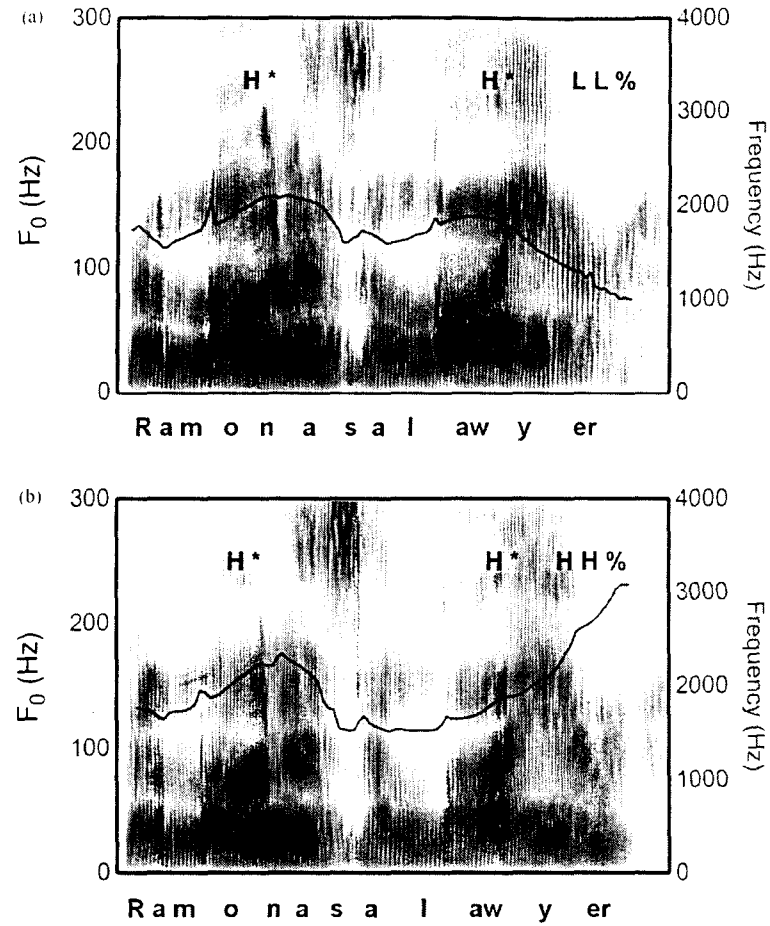


Figure 3.2. *The English sentence* Ramona's a lawyer, *spoken (a) as an echo question with the accent sequence H* ... H* ⊥H ⊥H%, and (b) as a neutral declarative with the accent sequence H* ... H* ⊥⊥%. In terms of a traditional British analysis, the echo question has a nuclear high rise and the neutral declarative a nuclear fall (or high fall).*

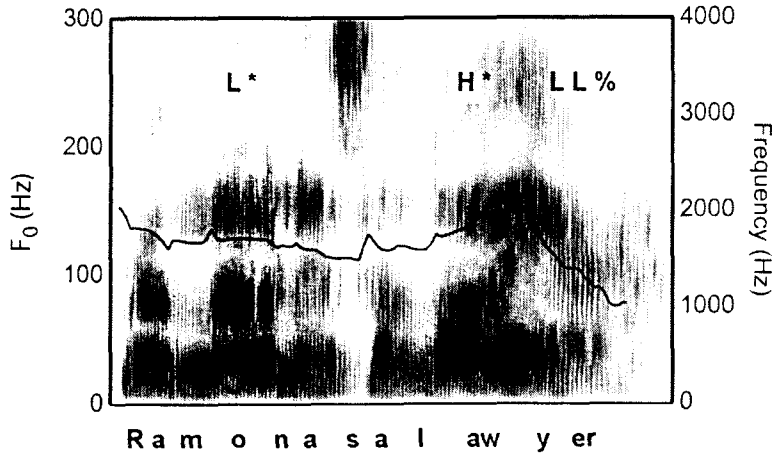


Figure 3.3. *The English sentence* Ramona's a lawyer, *spoken as a declarative sentence with the accent sequence L* ... H*⊥L%. The prenuclear low accent combined with the nuclear falling accent yields what Liberman 1975 referred to as the 'surprise-redundancy contour'.*

tradition (e.g. Kingdon 1958; O'Connor and Arnold 1973) called 'falling' and 'rising-falling' nuclear tones, or Tone 1 and Tone 5 respectively in Halliday (1967a). In these contours, the peak of H* is generally aligned with the accented vowel, while in L*+H the valley corresponding to the L* is aligned with the accented syllable and the peak is aligned later. Although the notational distinction between H* and L*+H is an obvious way to represent this difference in an AM analysis, there is actually a long-standing disagreement whether the falling and rising-falling nuclear tones should be regarded as involving distinct accent types. Some versions of the British tradition (in particular the original version proposed by Palmer 1922) make no reference to the rise–fall at all, so that the extent of any rise before the fall is implicitly treated as a matter of paralinguistic variation within the single category 'fall'. In an analysis influenced by the British tradition, Vanderslice and Ladefoged (1972: 822) proposed a compromise expressed in terms of distinctive features: they suggested that the rise–fall should be treated as a variant of the fall on the grounds that the difference between the variants 'clearly convey[s] indexical or paralinguistic information', and they analysed the phonological difference in terms of a feature 'Scoop'. This idea was imported into the AM tradition as a feature 'Delayed peak' (Ladd 1983a), or simply 'Delay' (Gussenhoven 1984), which is supposed to apply to the H tone or to the nuclear accent as a whole. This issue is still far

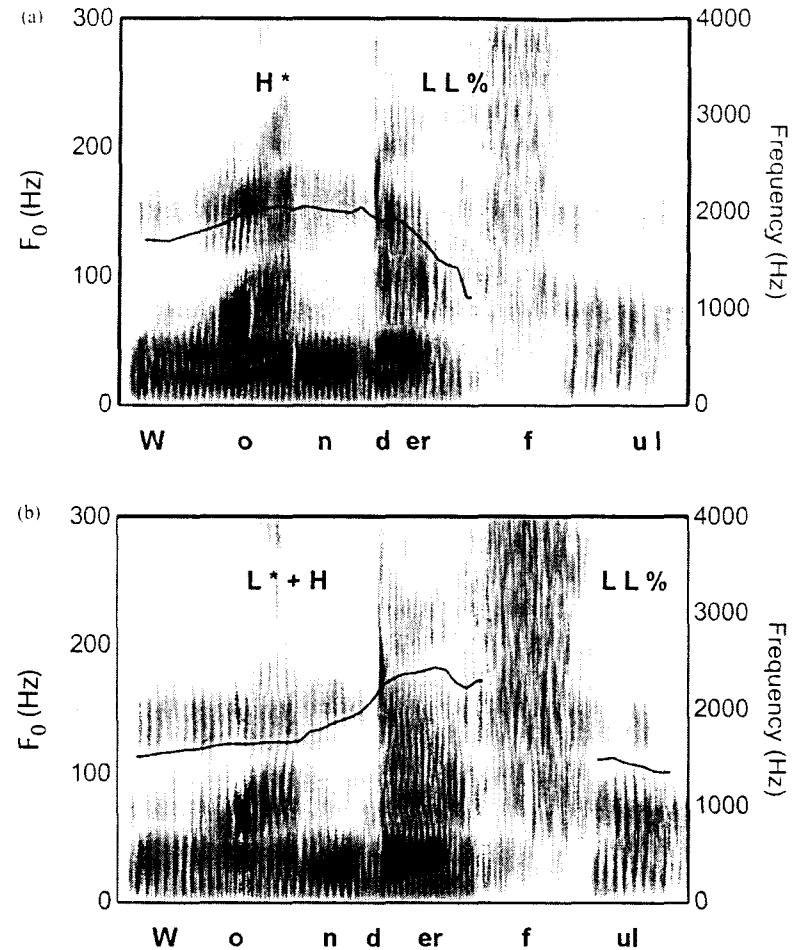


Figure 3.4. *The English word* Wonderful, *spoken (a) as a neutral declarative statement or citation form with the accent sequence H*⊥L%, and (b) as an emphatic statement with the accent sequence L*+H⊥L%. Some British-style analyses have treated the difference between these two contour types as the difference between (a) fall and (b) rise–fall; others have treated both as variants of a single type (e.g. (a) plain versus (b) 'scooped' fall, or fall with (a) non-delayed or (b) delayed peak).*

from settled, in part because of the apparent contradiction between regarding the difference as paralinguistic (which suggests gradient variability; cf. section 1.4) and analysing it in terms of a linguistic feature (which suggests a categorical difference, at least in the Ladd version). The Ladd/Gussenhoven analysis is also difficult to make phonetically explicit (cf. section 4.1.2). However, the Pierrehumbert analysis has no way of dealing with the intuition that the two contour types are in some way variants of the same basic category, which is really the root of the long-standing taxonomic disagreement. We return to this question in section 4.2.

We now consider the three accent types whose phonetic interpretation in Pierrehumbert's analysis is less obvious, namely L+H*, H*+L, and H+L*. The L+H* accent has been problematical from the outset. In theory, it is typically characterised by pitch movement from a preceding syllable, rather than (as is the case with L*+H) movement to a following one. This is exactly what the star notation suggests: the starred tone of both accent types is the defining one and is aligned with the accented syllable, while the unstarred tone is a leading tone in one case (L+H*) and a trailing tone in the other (L*+H). However, this description means that there is a problem distinguishing between L+H* and monotonal H*, because, as we saw in figures 3.2 and 3.3, H* may represent a local pitch jump on the accented syllable. This local pitch jump also implies pitch movement from a preceding syllable. According to the ToBI guidelines, 'the L+H* differs from the H* primarily by a *more substantial rising pitch movement* leading up to the H* target, i.e. the presence of a preceding L target' (Brugos, Shattuck-Hufnagel, and Veilleux 2006: sect. 2.5.2; emphasis added). The difference between L+H* and H* is therefore often fairly clear if there is a preceding syllable to display the level of the leading L, but on a phrase-initial accented syllable L+H* and H* are difficult to distinguish, and it is known that the distinction between these two accent types is hard for transcribers to make reliably (Pitrelli, Beckman, and Hirschberg 1994).

In a sense it is the phonetic interpretation of H*, not L+H*, that is the problem. The fact that H* is realised phonetically as a local pitch jump might seem to suggest the presence of a leading L before the H* in all cases. The supposed distinction between the H* and the L+H* therefore raises the more general question of how to identify which points in a contour reflect the occurrence of tonal targets. If both H* and L+H* are characterised by a rise in pitch at the beginning of the accented syllable, on what basis do we decide that the beginning of the rise is an L tone in one case and no tone at all in the other? This issue has been the topic of considerable discussion and empirical work

(cf. in particular Ladd and Schepman 2003), and Pierrehumbert's proposal that English has three distinct high accent types, H*, L+H*, and L*+H, remains problematical. We return to this issue at several points below.

The final two bitonal accent types – H*+L and H+L* – raise the question of the phonological treatment of downstep, and a short digression is required before we discuss them. As we saw in section 2.4, there is wide agreement with Pierrehumbert's idea that English involves a distinction between downstepped and non-downstepped accents, but there has been considerable disagreement about the relation between downstep and the tonal composition of specific accents. In Pierrehumbert's original analysis, downstep was said to be triggered (as it is in many African languages) by any H ... L ... H tonal sequence, regardless of what part the three tones play in pitch accents (i.e. sequences like H*+L ... H*, H* ... L+H*, L+H* ... L+H*, and so on would all trigger downstep of the second H tone). In the revised Beckman and Pierrehumbert analysis, the claim is modified: all bitonal pitch accents (viz. H*+L, H+L*, L*+H, L+H*) trigger downstep of the following *accent* (including phrase accents), regardless of whether there is a tonal sequence H ... L ... H. (This modification is what leads to the multiple interpretations of tone strings in table 3.1.) In both cases, downstep is seen as a matter of the phonetic realisation of specific *sequences* of tones or accents. The alternative view, first put forth in connection with AM analyses of English by Ladd (1983a), is that downstep is itself an independently selectable phonological option, which can apply, or not, to essentially any accent in any sequence. This independently selected downstep specification is indicated by an exclamation mark preceding the affected tone – a typographical simplification of the traditional Africanist notational device /ꜜ/.

The difference between these views has consequences for the description of pairs of utterances like the two versions of *my mother's diaries* in examples (21) and (22) in chapter 2, and this brings us to the phonetic realisation of the H*+L accent in Pierrehumbert's analysis. Panel (a) of figure 3.5 shows the non-downstepping version (21), while panel (b) shows the downstepping version (22). Both have two pitch accents; in the Pierrehumbert system, the distinction between the two sequences of accents is analysed as a function of the phonological nature of the *first* accent in the sequence, rather than the second. Specifically, the downstepping contour is analysed as a sequence H*+L ... H*, in which the H*+L accent is said to trigger downstep in the following H* accent, whereas the non-downstepping contour is analysed as a sequence of two H* accents. By contrast, in the view that developed from the

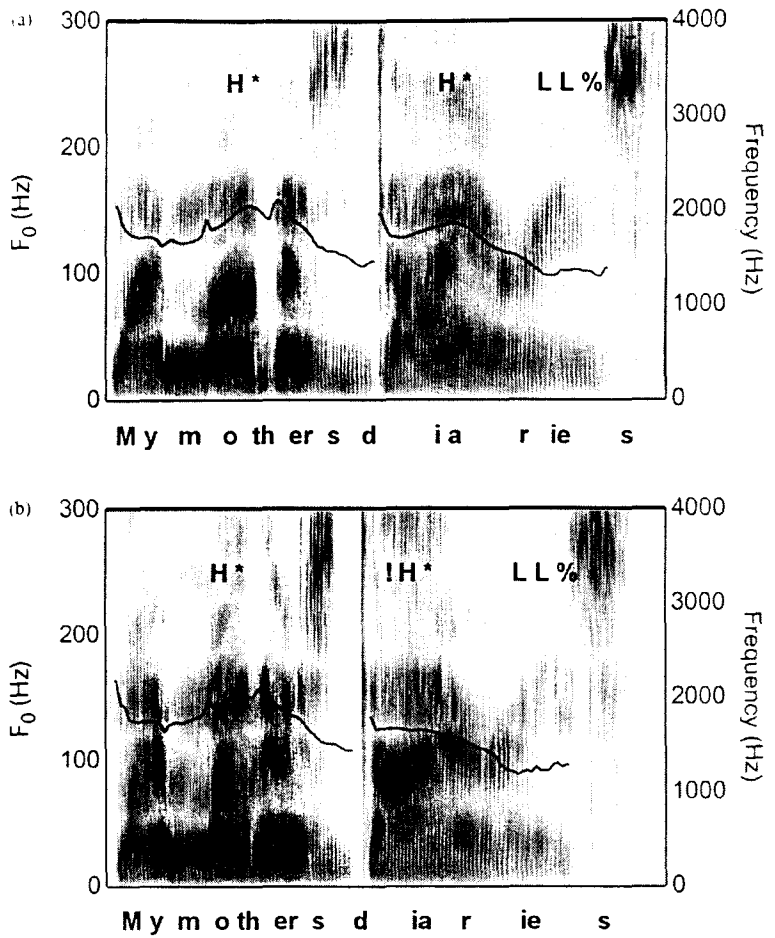


Figure 3.5. *The English phrase* My mother's diaries, *spoken (a) as a neutral or slightly emphatic statement or citation form with the accent sequence H* ... H*⊥⊥%; and (b) as a statement or citation form expressing some sort of finality and/or familiarity, with the downstepping accent sequence H* ... !H*⊥⊥%. For more detail see text.*

proposal in Ladd 1983a, both contours are seen as composed of two identical high accents; the difference is that in the downstepping contour the second high accent has an additional downstep feature that is largely independent of what precedes. The accent sequences would thus be more transparently notated H* ... !H* in the downstepping case, and H* ... H* in the non-downstepping

one. The more transparent notation has come into general use since the early 1990s.[5]

Before leaving the H*+L accent, we should point out that the notation H*+L has also been used in a very different way by a number of authors to refer to an accent with a peak followed by a fall. This usage was proposed for English, for example, by Ladd (1983a) and Gussenhoven (1984: ch. 6), and has been used by a number of other researchers for other languages (e.g. Féry 1993 for German; Frota 1998, 2002 for Portuguese; Gussenhoven 2005 for Dutch). In all cases, these proposals refer to a pitch pattern that could alternatively be analysed as an H* or L+H* accent followed by an L phrase accent, and the question of whether to use H*+L in this way is therefore basically about whether to accept the idea of the phrase accent. This issue will reappear repeatedly throughout this chapter, and again in section 4.1.4.

As for the H+L* accent, it transparently indicates a local movement from a higher preceding unaccented syllable to a lower accented one, and the relationship between H+L* and L* is thus somewhat comparable to that between L+H* and H*. However, the parallelism is not perfect, because in Pierrehumbert's usage the 'Low' target of H+L* is not a local valley or F_0 minimum. Instead, it is scaled at the level of a 'downstepped' or locally lowered *High* tone: that is, the pitch drop indicated by the H+L* notation is from one high pitch to another high pitch that is slightly lower, not from a high pitch to a level that would independently be described as low. The main use of this accent type in Pierrehumbert's original analysis is in describing 'terraced' downstepping contours in English, corresponding to the IPO 'Type E Fall'. An example is shown in figure 3.6. Although there are certain motivations for analysing this accent type as H+L* (see Pierrehumbert 1980: ch. 4), subsequent developments

5 Beckman and Pierrehumbert (1986) point out a potential problem with the idea of a 'downstep feature', namely that it could in principle be applied to the first accent in a series. They argue that the very notion of downstep implies reference to a preceding accent, and point out that the Pierrehumbert approach, which treats downstep as a property of accent *sequences*, is inherently unable to downstep the first accent in a series. This objection must certainly be taken seriously, but can be avoided if we interpret downstep as a kind of 'metrical' relation between accents, as discussed in section 8.3.3. Moreover, it is worth noting that it may not be entirely meaningless to talk about downstepping the first accent, so long as we assume that there is comparability of pitch level across utterances. That is, we might in principle find that, other things being equal, some utterance-initial H tones are systematically lower than others. In fact, both Connell and Ladd (1990) and Laniran (1992) independently looked for evidence of such 'initial downstep' in Yoruba – where it is relatively simple to compare the average pitch of initial H tone with that of the H of an initial L ... H sequence – and both failed to find any. However, in both cases it was easy to state what would constitute evidence for initial downstep and hence to test the idea empirically, which suggests that the idea of a downstep feature is not prima facie incoherent.

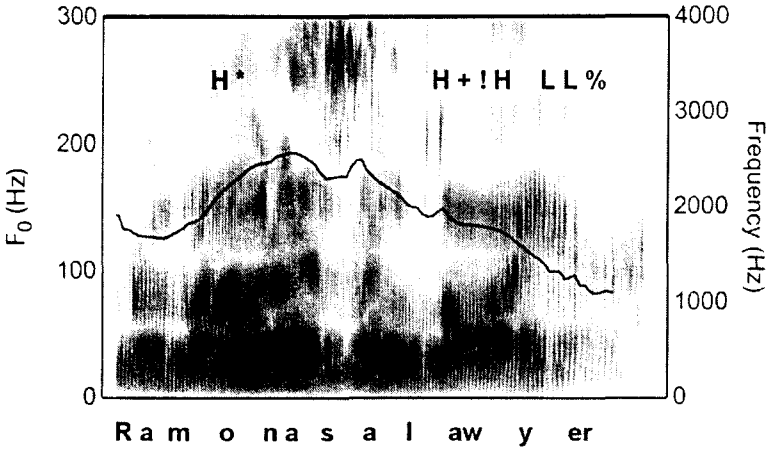


Figure 3.6. *The English sentence* Ramona's a lawyer, *spoken as a statement expressing some sort of finality, with the downstepping accent sequence H* ... H+!H* ⊥⊥%. The nuclear pitch accent H+!H* is used to indicate that the pitch stays relatively high following the prenuclear H*, and the downstep takes place immediately before the nuclear accent syllable.*

in AM usage have moved away from Pierrehumbert's original notation. Two distinct notational innovations are involved here. First, the accents in terraced downstepping sequences are now transcribed H+!H* in the ToBI standard, clearly indicating that the level of the accented syllable is not necessarily low in the speaker's range. Second, the H+L* notation is now widely used (e.g. Frota 2002 for Portuguese; D'Imperio 2002 and Grice *et al.* 2005 for Italian; Grice, Baumann, and Benzmüller 2005, 2007 for German) to indicate a pitch accent in which the accented syllable is not only preceded by a higher-pitched unaccented syllable (hence H+), but actually is more or less at the bottom of the speaker's range (hence L*). An example of this accent type from Italian is shown in figure 3.7.

3.1.4 Edge tones

'Edge tone', as noted above, is used here as a cover term for what Pierrehumbert referred to as phrase accents and boundary tones. As we saw in the preceding subsection, edge tones are in theory entirely independent of pitch accents, and the presentation here aims to make clear the motivations for the original theoretical claims; in particular, phrase accents and boundary tones are discussed separately. However, as implied by our notational conventions (cf. footnote 2), it seems perfectly coherent to recognise combinations of pitch accent and edge

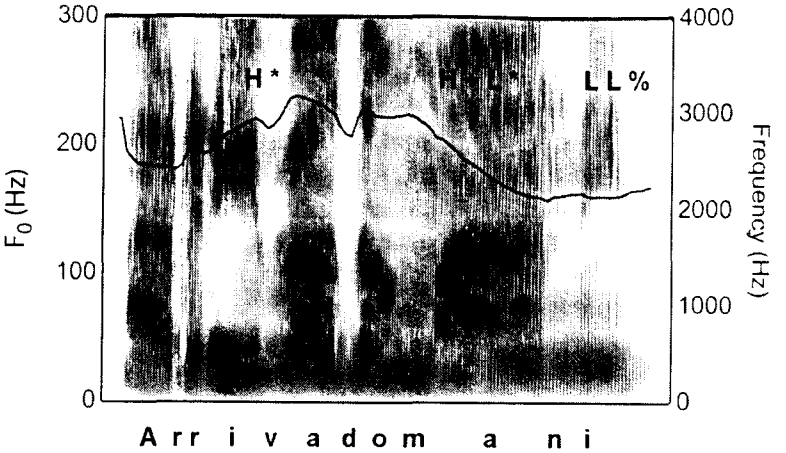


Figure 3.7. *The Italian sentence* Arriva domani *('S/he arrives tomorrow'), showing a downstepping contour similar to the English contour in figure 3.6, and illustrating the common analysis of the nuclear accent as H+L*. See text for more detail.*

tone(s) as some kind of higher-level unit like a British-school 'nuclear tone', and the degree of real independence of pitch accent and edge tone has long been an unresolved issue within AM theory.

3.1.4.1 Phrase accents

The phrase accent idea is borrowed from Bruce's (1977) analysis of Swedish. As we saw in section 1.3.3, Bruce showed that a Stockholm Swedish intonation contour can be analysed as a sequence of one or more 'word accents' (the lexically specified accent patterns), followed by an additional peak at the end of the focused word of each phrase, followed at the end of the utterance by a fall to the bottom of the speaker's range. The peak at the end of the focused word of each phrase is the 'phrase accent'.[6] What Pierrehumbert proposed was that in English, as in Swedish, the last pitch accent of each phrase is followed by two distinct tonal events. This analysis is motivated most clearly in the case of the rising–falling–rising nuclear accent contour of English, where the first rise is analysed as an L+H* or L*+H pitch accent; the fall is the reflex of an L phrase accent; and the final rise reflects an H% boundary tone. In many other

6 Bruce actually used the term 'sentence accent', but Pierrehumbert revises this without comment to 'phrase accent'. This revision can readily be justified by the already existing profusion of meanings for the term 'sentence accent' in English.

cases the phrase accent analysis is not so clearly motivated, and as we just saw there have been a number of suggestions that a nuclear falling contour should be analysed as an H*+L pitch accent rather than as a sequence of an H* pitch accent and an L phrase accent. There are a number of AM descriptions (perhaps most notably the ToDI labelling system for Dutch; Gussenhoven *et al.* 2003; Gussenhoven 2005) which do not make use of the phrase accent at all.

The details of how the phrase accents are used in Pierrehumbert's analysis of English are as follows. The L phrase accent can be straightforwardly interpreted as low pitch following the final pitch accent of the phrase. The fall to this low level takes place fairly soon after H* or any bitonal pitch accent, and the pitch remains low until the final boundary tone, at which point it either rises abruptly (H%) or drops off to the bottom of the speaker's range (L%). The stretch of low pitch can span several syllables, as in a sentence like (1):

(1)

We TOLD you we weren't coming!

This extended span of low pitch is described by Pierrehumbert (1980: ch. 5) in terms of a 'tone-spreading' rule of the sort found in many African tonal systems, but there are other possibilities. For example, Grice, Ladd, and Arvaniti (2000) suggest that the phrase accent may be multiply associated – to both the nuclear accented syllable and the end of the phrase – and the extended span of low pitch is simply an interpolation between the first low target and the second; Lickley, Schepman, and Ladd's findings (2005) could be taken as evidence for this view.

The H phrase accent should analogously represent a high pitch following the last pitch accent, but in some cases the empirical basis of this description is problematical. The main problem is with tunes that begin rising at the last accent and continue to rise until the end, such as the patterns that the British tradition would treat as high-rise and low-rise nuclear tones. In the sequences H*_H_H% and L*_H_H% there is no obvious phonetic basis for saying where in the contour the H phrase accent 'occurs'; similarly, in a 'low-rising' sequence there is no obvious basis a priori for choosing between the description L*_H_H% and L*+H_H_H%.[7] However, as we shall see below (section 3.2.3.1), there is at least one pitch pattern in English that provides clear justification for the idea

7 This statement is true only for the original Pierrehumbert 1980 version of the analysis. In the Beckman and Pierrehumbert 1986 version, the modified analysis of downstep means that these two sequences are realised differently. Cf. table 3.1.

of a well-localised H phrase accent, namely the 'calling contour' (what Ladd 1978 called the 'stylised fall'), as in:

(2)

Jennifer!

The step down in pitch indicates the location of the (downstepped) H phrase accent.

3.1.4.2 Boundary tones

The description of boundary tones in Pierrehumbert's analysis of English is as follows. The H% boundary tone always indicates a final rise, which typically takes place at the very end of the phrase or utterance. This is unproblematic after the L phrase accent, but requires an 'upstep rule' (Pierrehumbert 1980: sect. 4.5) after the H phrase accent. That is, the final rise is the phonetic outcome one would expect from a sequence of L and H, but from a sequence of H and H one might expect only a sustained level, and the upstep rule ensures that the level of the H% boundary tone is higher than that of the immediately preceding H phrase accent. In both cases, H% is clearly to be identified with the Type 2 Rise of the IPO analysis of Dutch, which is a 'non-prominence-lending' pitch movement occurring at the ends of phrases. As noted in section 1.2.1, Cohen and 't Hart (1967) originally found it odd that a pitch movement could be so distinctive and so unambiguously localised and yet not be prominence-lending, but it now seems beyond dispute that the recognition of discrete pitch movements at boundaries is a concrete advance in our understanding.[8]

The L% boundary tone, as Pierrehumbert proposes to use it, can best be described as indicating not a final fall but merely the absence of final rise. After an L phrase accent, it indicates a gradual continued fall to the bottom of the speaking range, but after an H phrase accent it indicates a sustained level pitch, contrasting with the additional final rise represented by H%. That is, the sequence H_L% in Pierrehumbert's notation means 'sustained level phrase-final pitch'. Except in the case of the calling contour, which appears to be part of a more general phenomenon of 'stylised intonation' (Ladd 1978), sustained level pitch is not very common in English, but it is frequent in both German and Dutch. In describing English stylised intonation, I have elsewhere (Ladd

8 In the British tradition, there is no analogue to H%, and as a result British nuclear tones such as fall–rise conflate the pitch movement on the last accent (e.g. fall) with the pitch movement at the end of the phrase (e.g. rise). Because of this, as noted in section 2.1, the taxonomy of falling–rising contours is a long-standing unresolved problem in the British tradition.

1983a) proposed that the sustained pitch indicates the absence of any boundary tone. This transcriptional device has since been adopted by ToDI (Gussenhoven *et al.* 2003) for Dutch. More or less equivalently, Grabe 1998, in her analysis of German intonation, has proposed that sustained level phrase-final pitch should be represented more transparently by means of a boundary tone 0%, distinct from both H% and L%. Arvaniti and Baltazani 2005 propose to distinguish three boundary tones in Greek – H%, !H%, and L% – at least partly in order to be able to transcribe sustained level final pitch. We return to this issue in section 3.2.

3.1.5 ToBI

In the early 1990s, Pierrehumbert's analysis of English was adapted for use in a proposed standard prosodic labelling system for English speech databases known as ToBI (Tones and Break Indices) (Silverman *et al.* 1992; Pitrelli, Beckman, and Hirschberg 1994; Beckman and Ayers Elam 1997; Brugos, Shattuck-Hufnagel, and Veilleux 2006). A complete ToBI transcription actually contains several 'tiers' – strings of symbols anchored in time to specific points in the waveform of an utterance – including the orthographic transcription, a tier reserved for comments on disfluencies and the like, and a tier used for suggestions of alternative possible transcriptions. Of these multiple tiers, the two most important are those indicating the *tones* in the F_0 contour and the *break indices* that label the strength of each word boundary; and these two are what give the system its name. However, in a great many contexts, phrases like 'ToBI labels' and 'ToBI transcription' refer only to the tonal tier. For example, as early as the mid 1990s, published research on intonation with no connection to labelled speech databases and no interest in boundary strength began to make use of the tonal symbols of the ToBI transcription standard wherever necessary to describe pitch contours (e.g. Birch and Clifton 1995). Consequently, no discussion of AM intonational phonology would be complete without a presentation of the ToBI notational conventions and the numerous extensions of ToBI to other languages.

In order to provide some context for understanding the modifications that were made to the original Pierrehumbert analysis when it was adapted for ToBI, it is worth briefly sketching the background to the development of the system. ToBI was the joint initiative of a group of American researchers whose primary interest was in establishing a common system to indicate prosodic features in labelled computer corpora of speech. Two of the principal groups involved in this initiative had developed their own partial labelling systems. Those we may refer to as the 'To' group were more interested in intonation in the narrow sense of utterance melody, and had worked with Pierrehumbert's tonal analysis of English. The 'BI' group were more concerned with phrasing, prominence, and prosodic structure generally, and had developed the notion of 'break index' as a way of impressionistically indicating the strength of word and phrase boundaries (e.g. Price *et al.* 1991). The development of the system involved not only compromises between these two main groups, but also detailed discussion with potential users of labelled corpora (for research in speech technology or the study of dialogue). Once a preliminary system was in place, there was further discussion with representatives of certain other schools of prosodic transcription (in particular IPO and traditional British) before the agreed standard was finally published in 1992. There have been a few modifications since then, perhaps the most conspicuous of which is that the system is now officially referred to as MAE_ToBI (Mainstream American English ToBI) (Beckman, Hirschberg, and Shattuck-Hufnagel 2005), to take account of the fact that ToBI standards have more recently been developed for several languages other than English.

The tonal tier of MAE_ToBI uses five pitch accent types (H*, L*, L+H*, L*+H, and H+!H*); two phrase accents (H- and L-); two final boundary tones (H% and L%); and one initial boundary tone (%H). In addition, it uses the downstep diacritic /!/, which can be applied to any H tone (except H%) that is preceded by another H tone in the same phrase (i.e. !H*, L+!H*, L*+!H, and !H- are all possible, but not as the first accent in a phrase). The system also allows the transcriber to express uncertainty about the type of a pitch accent (X*) and about whether a pitch accent is present or not (*?). The latter type of uncertainty is discussed in more detail in section 7.1.3.

The break index tier labels every orthographic word boundary with a number from 0 to 4. Break indices 3 and 4 correspond to intermediate phrase and intonation phrase boundaries, as discussed in section 3.1.1 above. Break index 1 represents the normal expected disjuncture between words in sequence, and break index 0 indicates some special closeness (e.g. the presence of a *tch* transition between *get* and *your* in a sentence like *Did you get your ticket?*). Break index 2 is used to signal a mismatch between tonal and other cues to disjuncture between words; this will be discussed shortly.

The reduction of Pierrehumbert's six pitch accent types to five, and the adoption of the downstep diacritic, were based on a compromise in the development of the standard between those who wanted phonetically transparent labels and those who were comfortable with the idea of complex phonetic realisation rules working on an abstract phonological analysis. Specifically, the agreed standard abandons Pierrehumbert's idea that downstep is a matter of the phonetic

realisation of specific accent *sequences*, and adopts the idea proposed by Ladd 1983a that downstep is a meaningful modification of individual accents that can apply at least somewhat independently of the tonal context. This change was made partly because in actual auditory transcription it is not always clear whether an accent is to be regarded as downstepped (cf. Dainora 2001), and transcribing the distinction with a diacritic rather than a different accent type somehow reflects this uncertainty. More importantly, the change was made in part because some of the contributors to the system found the standard Pierrehumbert notation difficult to use even when there was no doubt about the presence of downstep. Two specific modifications to Pierrehumbert's system resulted. First, the H*+L was eliminated from the inventory of accent types, because, as we saw in section 3.1.3, it is only ever used as a downstep trigger. Any accent that would be transcribed H*+L in Pierrehumbert's system is transcribed H* in ToBI, and the succeeding accent is marked with the downstep diacritic /!/. Second, the H+L* notation, which in the standard Pierrehumbert system indicates a downstepped accent that steps down locally from the preceding unstressed syllable, was replaced by the notation H+!H*. As noted in section 3.1.3, this new notation makes it clear that the level of the accented syllable is not necessarily low in the speaker's range, but only lower than what precedes, and in some sense still counts as 'High'.

The other important compromise in the development of the agreed standard involved the reduction of the break index scale from seven points to five, and the acceptance of the notion of the phrase accent, although the latter was still a matter for major disagreement when the standard was adopted in the early 1990s. Specifically, the standard adopted Beckman and Pierrehumbert's idea that the phrase accent is the edge tone for an intermediate phrase, which means that any transcription on the tonal tier automatically implies three possible degrees of disjuncture between words, namely intonation phrase boundary (T%), intermediate phrase boundary (T-), and no phrase boundary. Now, the original break index scale (Price *et al.* 1991) was intended for purely impressionistic transcription, and there is experimental evidence (de Pijper and Sanderman 1995) that listeners are fairly reliable in their estimates of the degree of disjuncture between words. Nevertheless, the agreed ToBI standard effectively subordinates the break index tier to the tonal tier, and strict constraints were put on the impressionistic use of break indices to mark degrees of disjuncture. Break indices 3 and 4 take on the specific meanings of *ip* break and *IP* break respectively, and any well-formed transcription that shows a phrase accent on a given word must therefore normally also show a 3 break index between that word and the following word.

This leads to problems in transcribing F_0 contours which, according to the Pierrehumbert analysis, can arise only through a sequence of pitch accents and phrase accents, but which do not contain an audible boundary or disjuncture. For example:

(3)

Edinburgh's the capital of Scotland.

This must be analysed in Pierrehumbert's system as involving a phrase accent and hence an intermediate phrase boundary:

(4) (a) H* L H* LL%
Edinburgh's the capital of Scotland.

Yet it is perfectly possible for there to be no cues to the presence of a boundary between *Edinburgh's* and *the* except for the phrase accent itself. To those unconvinced of the very validity of the phrase accent analysis – like some of those who participated in the discussions that led to the ToBI standard – the proposed requirement to coordinate a break index 3 with the phrase accent seemed circular at best, and at worst misleading about the phonetic reality that the labelling was intended to reflect. The problem was solved in ToBI by permitting break index 2 to be used in conjunction with a phrase accent on the tonal tier in cases precisely like these. That is, break index 2 can indicate the absence of cues to the presence of a boundary other than those required by the tonal analysis. Thus:

(4) (b) H* L H* LL%
Edinburgh's the capital of Scotland.
2 1 1 1 4

Break index 2 is also used in the converse way, to indicate an audible break between words that has no apparent tonal correlates. This compromise allowed the discussions to go forward, but at the expense of making incomplete use of transcribers' impressionistic judgements of disjuncture.

3.2 AM descriptions of intonation in other languages

3.2.1 Beyond English: AM theory and AM analyses

Well before the advent of ToBI, Pierrehumbert's analysis of English had quickly established itself as the norm against which other descriptions of intonation should be measured, and the autosegmental–metrical theory of intonational structure that underlies it soon began to influence work on other languages. My

own early response (Ladd 1983a), though in some respects a critique, showed how the theory could be used to understand the phenomena of Hungarian question intonation (cf. section 2.5), and to overcome certain difficulties in the IPO analysis of Dutch. Further AM work on Dutch was done about the same time by Gussenhoven and his colleagues (Gussenhoven 1984; van den Berg, Gussenhoven, and Rietveld 1992; Rietveld and Gussenhoven 1995). A group of coordinated projects on intonation in Germany during the 1980s led to early adoption of AM ideas in the description of German (e.g. Wunderlich 1988; Uhmann 1991; Féry 1993), while Hayes and Lahiri (1991) provided a succinct and elegant application of the theory to the intonation of Bengali. By the mid 1990s, a number of articles and dissertations had already appeared that adopted the general AM view in dealing with the details of specific languages or dialects, including Grice 1995a on Palermo Italian; Post 2000 (developed from a 1993 MA thesis) on French; King 1998 (written 1992) on Dyirbal; Sosa 1999 (based on a 1991 PhD thesis) on Spanish; and Jun 1993 on Korean.

The pace of this research has quickened enormously since the mid 1990s. First of all, there are now reasonably well-developed ToBI-style transcription systems for as many as a dozen languages, including Dutch, German, Greek, Italian, Japanese, and Korean (see the papers in Jun 2005 for good summaries of these and other AM descriptions). It would be foolish to attempt to give a detailed guide to this work, as it would be out of date before it appeared; the main ToBI website (www.tobihome.org) is the best way into this flourishing field. Second, a great deal of other AM work not closely integrated with the ToBI movement, and not specifically aimed at providing a labelling system for speech databases, has contributed to the expansion of our knowledge on several other well-studied languages, notably the Romance languages (e.g. D'Imperio 2002; Frota 1998, 2002; Prieto, D'Imperio and Gili Fivela 2005; Prieto and Torreira 2007). Finally, still other new work extends the AM theory to languages whose prosodic systems were previously completely unstudied (e.g. Rialland and Robert 2001 on Wolof; Bishop and Fletcher 2005 on Bininj Gun-wok (Mayali); Gordon 2005 on Chickasaw; Lindström and Remijsen 2005 on Kuot; Remijsen and van Heuven 2005 on Papiamentu; and Himmelmann unpublished ms on Waima'a). Though one cannot entirely discount intellectual fashion as a motivation for this rush to the AM view, I think it is fair to say that part of the reason for its rapid acceptance is that it genuinely sheds new light on old problems in the description of intonation.

Nevertheless, even the limited discussion in this chapter so far should make clear that there is not a single 'correct' AM analysis of the intonation of English or any other language. Rather, what the AM approach to intonational phonology

provides is a framework of expectations about the kinds of structures we may find in the intonational system of a new language. Within this framework of expectations, many questions remain to be resolved – for example, the usefulness of the phrase accent construct – and individual investigators have answered these questions in different ways while still, in some sense, adopting the overall body of AM assumptions. This has been my principal motivation for distinguishing the presentation of fundamental tenets of the AM theory in the previous chapter from that of the specific descriptive conventions in this chapter. It is important to see Pierrehumbert's analysis of English not as a unitary theory, but as a specific proposal for the description of English intonational phonology within the framework of a broader set of theoretical assumptions.

It is especially important to keep this in mind because of the potential value of the AM theory in facilitating cross-variety and cross-language comparison. We already saw in section 2.5 that the AM description makes the relation between pitch and focus in Hungarian question intonation appear ordinary rather than mysterious – but it does this only if we accept that Hungarian question intonation involves a different tune from question intonation in English or French, and a different neutral placement of the nuclear accent. Yet the value of such cross-language comparisons is only as good as the comparability of the descriptions. If we take different AM analyses at face value, we may forget that the AM theory involves important points of disagreement within its overall set of assumptions. Only by exploring the application of AM concepts to a variety of different languages will we be in a position to resolve the broader issues.

Take again the example of the phrase accent. If we compare published descriptions of Dutch (e.g. Gussenhoven 2005) and Portuguese (e.g. Frota 2002), which use H*+L for a falling nuclear accent and do not make use of the phrase accent at all, with Pierrehumbert's description of English (or Grice, Baumann, and Benzmüller's of German, or Arvaniti and Baltazani's of Greek), it probably does not make sense to conclude that Dutch and Portuguese do not have a phrase accent while English and German and Greek do. Rather, it is much more likely that nuclear or phrase-final accents in all these languages are structured in much the same way, and that we do not yet know for sure how best to describe them. Frota's discussion (2002) of this issue shows clearly that she takes this as a general theoretical question about the composition of nuclear accents cross-linguistically, not as a matter of a difference between Portuguese and English. Grice, Baumann, and Benzmüller (2005: 56) also clearly acknowledge that differences between different AM analyses of German 'are of a theoretical rather than a typological nature'.

This might seem like an obvious conclusion, but we can see that it is not if we consider similar differences over the analysis of sustained level phrase-final pitch in different languages. As we saw above, Pierrehumbert analyses sustained level phrase-final pitch in English as a sequence H_L%, contrasting with the 'upstepping' sequence H_H% that represents rising phrase-final pitch. Grice, Baumann, and Benzmüller (2005) have a similar analysis for German, except that they explicitly mark the upstep of the boundary tone in the final rise as ^H%. Various other writers (Ladd 1983a on English; Gussenhoven 1984, 2005 on English and Dutch; Grabe 1998 on English and German; Gordon 2005 on Chickasaw) have proposed that sustained level phrase-final pitch should be analysed as the absence of a boundary tone. Arvaniti and Baltazani (2005) allow for apparent contrasts of pitch height in sustained level phrase-final pitch in Greek by proposing a !H% boundary tone alongside H% and L%, and representing sustained level phrase-final pitch with a sequence of identical phrase accent and boundary tone (e.g. !H !H%). With so many different analyses of what is, on the face of it, a similar intonational phenomenon in different languages, it would seem appropriate to consider again the possibility that the different analyses are based on theoretical rather than typological differences. But Jun (2005: 448) seems to conclude that, for example, ^H% is specific to German rather than to the GToBI analysis of German. Similarly, Arvaniti and Baltazani (2005: 94) simply accept that upstep plays a role in the scaling of German and English boundary tones, and suggest that their analysis of sustained level phrase-final pitch in Greek is different because Greek is different. It is, of course, perfectly possible that Greek and English are phonologically different, but that alone does not mean that H_L% is the best analysis of sustained level phrase-final pitch in English. Eventually, that is, we will have to discuss divergent analyses for what they are, rather than taking cover behind the notion that every language is to be analysed in its own terms.

3.2.2 *Issues in cross-language comparison*

The problem of ensuring the comparability of descriptions is actually far from trivial, and is closely related to one of the issues that caused the greatest difficulty in the development of the original ToBI standard. During the discussions that led to the formulation of the standard, there was a good deal of disagreement between those who insisted that only phonologically distinct elements of the intonational system should be indicated, and those who felt that the system ought to allow for the transcription of impressionistically audible phonetic detail. One group – referred to here for want of a better label as the 'phonetic' group – pointed to the uncertain phonological status of certain features

that were deemed worthy of transcription (such as the distinction between H* and L+H*), and to the possibility that later developments would reveal that certain other details, left untranscribed in the agreed system, were phonologically important. The other group – referred to here with an equally unrevealing label as the 'phonological' group – insisted that because ToBI is a system for labelling speech files, the phonetic detail is always available for later inspection in the signal, and that a rigorous phonological analysis is the only defensible basis for deciding which elements to include in a set of symbolic labels.

More fundamentally, the 'phonological' group based their approach on an explicit theoretical understanding of symbolic transcription, namely Pierrehumbert's argument (1990; cf. Pierrehumbert 1980: sect. 1.1; Pierrehumbert and Beckman 1988: ch. 1; Pierrehumbert, Beckman, and Ladd 2000) that there is no scientifically valid 'systematic phonetic' level of description, only language-specific phonological abstractions that can and should be mapped directly onto the acoustic signal. This view sees any symbolic transcription (of any aspect of phonetic structure, not just of pitch) as ultimately based on some sort of analysis. Consequently, it rejects the idea of a theory-neutral symbolic notation for pitch (like INTSINT – Hirst and Di Cristo 1998: ch. 1; Hirst, DiCristo and Espesser 2000). It also – perhaps especially – rejects the idea (proposed by Nolan and Grabe 1997 and implemented in their IViE (Intonational Variation in English) project (Grabe 2003; Slater 2007); cf. also Hirst, Di Cristo, and Espesser 2000) that it is useful to distinguish between a syllable-based phonetic transcription of pitch and a more abstract phonological analysis. From the beginning, therefore, the developers of ToBI were at pains to point out that it is a *phonological* labelling system based on a specific *analysis* of English intonation. It is not, as many people were tempted to think, intended as a kind of 'hi-tech IPA' for intonation generally.

I reiterated this view in the first edition of this book (Ladd 1996: 95), and it is repeated prominently in ToBI materials on the web and elsewhere. For example, on the original ToBI home page (Beckman 2007), the following disclaimer is set out in its own text box: 'Note: ToBI is *not* an International Phonetic Alphabet for prosody. Because intonation and prosodic organization differ from language to language, and often from dialect to dialect within a language, there are many different ToBI systems, each one specific to a language variety and the community of researchers working on that language variety.' The point is taken up, forcefully and at some length, by Beckman, Hirschberg and Shattuck-Hufnagel (2005: 36–43). They quote Pierrehumbert's view (2000: 26) that a ToBI system 'is at the level of abstraction of a broad phonemic transcription', and that as such it 'must be guided by an inventory of [the language's] prosodic

and intonation patterns', and go on to say: 'This point cannot be emphasized too strongly' (Beckman, Hirschberg, and Shattuck-Hufnagel 2005: 36). While the theoretical question is certainly important, I believe that the situation is considerably less clear than Beckman *et al.* suggest.

There are two issues here: the informal 'semantics' of the symbols used in transcription and the problematic status of the abstractions involved in phonological analysis. With regard to the first of these, it is clear in theory that the choice of symbols for practical phonemic transcription of any given language is arbitrary. For example, it is of no real consequence whether we transcribe English *beat, bit, bait, bet, bat* as /biyt, bit, beyt, bet, bæt/ (e.g. Trager and Smith 1951), or as /bit, bɪt, bet, bɛt, bat/ (e.g. Giegerich 1992), or as /biːt, bɪt, beit, bet, bæt/ (e.g. Cruttenden 1994b). The differences between the transcriptions reflect mostly differences in the phonological analysis of the English vowel system, certainly not any consistent phonetic difference – in fact, phonetic details of different varieties (such as the great phonetic similarity between Australian *bat* and American *bet*) are for most purposes appropriately ignored. Anxious phonetics students who want to know which transcription is 'correct' need to be convinced that the choices are ultimately a matter of phonological analysis, not of what their ears can distinguish. Yet the IPA does provide a very approximate 'semantics' for its symbol set, so that while the transcription /bit/ can defensibly be used for the wide range of vowels found in the pronunciations of both the English words *beat* and *bit*, no one would suggest using /bit/ to transcribe *boot* or *boat*. The phonetic interpretation of any given IPA symbol is variable, but not limitlessly so. It is these informal limits that the 'phonetic' group took seriously when it argued that intonational distinctions involving downstep should be represented in a more 'phonetically transparent' way.

The second issue – the nature of phonological abstractions – is even more difficult. In the example of English vowels just cited, the main issue in comparing across varieties of English is phonetic realisation; phonologically, the set *beat, bit, bait, bet, bat* is comparable in all varieties. But this is not always true. For example, Scottish Standard English lacks any distinction between the vowels of *good* and *food*, and furthermore the vowel that is used in these words is typically quite short and rather fronted; a common IPA transcription uses the symbol [ʉ]. Do we identify this sound with the general English phoneme /u/? With /ʊ/? With neither? Our answer to this question will determine whether we say that Scottish English 'doesn't have /u/' or that it 'doesn't have /ʊ/' – or that it doesn't have either of them. We are on shaky theoretical ground, whatever we say. We are caught between the intuitive categorisation of sound types on which IPA transcription is based, and the enduring lack of theoretical clarity about the nature of phonemic categories. Indeed, these problems are familiar from many areas of linguistic description, not just phonetics and phonology: terms like *subject, noun, accusative, word*, and many others have approximate denotations that can usefully be applied to the description of many languages, yet seem to raise questions of detail in any individual language. There is a long-standing tension between acknowledging broad similarities across languages and recognising that, as a matter of strict theoretical interpretation, the categories of any language or dialect are part of a network of relations that make the categories unique to that system.[9]

However, there are unquestionable benefits to being able to compare across languages and dialects, and, as we saw in the case of Hungarian question intonation, the AM approach seems to offer new possibilities for making more insightful and more empirically testable comparisons. The rapid proliferation of AM analyses and complete ToBI systems for other languages besides English shows that ultimately this is something that many linguists want to be able to do. But in order to do that we have to accept the 'IPA paradox' – the fact that universally valid categories like [f] or [o], or even 'high back vowel', are difficult to define formally, and that theoretically rigorous definition of the categories of any specific language or dialect probably precludes theoretically rigorous comparison. In fact, the very idea that ToBI is 'a general framework for the development of prosodic annotation systems in other varieties of English ... and in other languages' (Beckman, Hirschberg, and Shattuck-Hufnagel 2005: 9) begs difficult conceptual questions about what it means to apply a common approach to transcription in different languages.

By far the most serious problem in cross-language comparison is the difficulty of establishing what we mean by 'the same' pattern in two different languages or different varieties of the same language, and the difficulty of establishing what we mean by saying that a given pattern 'does not occur' in a given language or language variety. A good illustration of these problems is provided by the falling–rising contour type used in yes–no questions, especially at the beginning of a transaction or a new topic – questions like *Could I have the*

9 See Lass 1984 for an interesting treatment of this theoretical problem as it relates to the description of vowel inventories; Bloomfield 1933: 84f. for an early statement of the theoretically awkward status of phonetic transcription generally; and Pierrehumbert, Beckman and Ladd 2000 for a discussion of the difficulty of squaring notions of phonological universals with scientifically sound phonetic descriptions of specific languages. This issue has more recently been taken up in an exchange between Haspelmath (2007) and Newmeyer (2007) on the possibility and desirability of positing language-independent categories for use in making typological generalisations of any sort, not just in phonology and phonetics.

bill please or *Is your mother there?* On sentences like these, it is quite common in Southern British English and many varieties of Dutch and German, but not in any North American variety of English, to use the tune shown, which we might analyse as a high H* prenuclear accent followed by a nuclear H*_L_H%.

(5) (a) H* H* LH%
Could I have the bill please

(b) H* LH%
Is your mother there?

In North American English such questions are most likely to have a high-rising nucleus H*_H_H%. The Southern British tune tends to sound condescending or peremptory to North Americans, since in North American English, to the extent that it is used at all, it is largely restricted to use on echo questions without question syntax implying a strong measure of doubt or disbelief:[10]

(6) H* H*LH%
You bought a Mercedes?

Given this background, it is interesting to consider the reaction of a native bilingual speaker of German and American English of my acquaintance to a discussion of this tune. In many varieties of German, as just noted, the H* ... H*_L_H% tune is normal with questions like those in (5):

(7) (a) H* H* LH%
Haben Sie heute Weizenbrot? 'Have you got wheat bread today?'

(b) H* LH%
Ist deine Mutter da? 'Is your mother there?'

The bilingual speaker shared the general American reaction when the tune was applied to English questions like (5), and was astonished to realise that exactly the same tune applied to comparable questions in German does not have the same force.

These comparative statements about German and American English seem accurate, valid and potentially interesting, and, like the case of Hungarian question intonation discussed at the end of chapter 2, they seem to point to the potential value of the AM approach in advancing our understanding of

10 Given how distinctively British this contour is, it is remarkable to find little discussion of it in the traditional British literature on intonation. Halliday, the only one who unambiguously treats it, calls it a 'Tone 2', which he distinguishes both from a 'Tone 2' (the ordinary rising question tune, L*_H_H% or H*_H_H%), and from 'Tone 4' (the English (rise–)fall–rise contour, L*+H_L_H% or L+H*_L_H%). In the rest of the British tradition, the contour under discussion is not clearly distinguished from the non-interrogative fall–rise.

intonation across languages. Just as it appears useful and meaningful to say that Hungarian question intonation is different in specific ways from question intonation in English or French – that is, to say that the same meaning is conveyed by different tunes in different languages – so in the present case it appears useful to be able to say that the same tune can be used in German and British English, or that the same tune conveys different meanings in British and North American English. Yet strictly speaking, the phrase 'the same tune' has no clear interpretation. Since the contours differ in detail depending on the text to which they are applied, the statement that all these utterances have 'the same' tune implies some sort of analysis; since the most obvious theoretical basis for such an analysis is a language-specific phonological inventory of tune types, the idea that utterances *in different languages* have 'the same' tune can easily be dismissed as meaningless. If we insist that our transcriptions are language-specific, we are left with no theoretically meaningful way to pursue cross-language comparison.

I believe that the AM approach to intonation provides us with a similar paradox. By making cross-language or cross-variety comparisons of the meaning of sequences like H* ... H*_L_H%, as I have just done, we skirt the same problems of strict theoretical interpretation and reap the same descriptive advantages that we derive from an informal metalanguage like the IPA, or from the application of terms like *subject* and *accusative* to a wide range of grammatical systems. With this in mind, the next section presents three interrelated case studies illustrating in some detail the potential of the AM approach for understanding cross-variety and cross-language differences of intonation, *so long as we allow things to be considered 'the same' across languages and across varieties.* A secondary purpose is to show once again that 'the AM approach' should not be equated with 'Pierrehumbert's analysis of English'; within the general framework of AM assumptions discussed in chapter 2, there are different possible analyses of various intonational phenomena.

3.2.3 *Intonational phonology across languages: case studies*

Before beginning, it is important to draw attention to the fact that, in intonation as in segmental variation, there are 'different types of differences' that it may be possible and useful to classify. The basic idea is set forth clearly by Wells (1982), who, in comparing the segmental phonetic aspects of different varieties of English, distinguishes *lexical–incidential*, *systemic*, *realisational*, and *distributional* differences. *Lexical–incidential* differences are essentially unsystematic differences in the choice of phonemes in specific words: for

example, *methane* has the vowel of *meet* in British English but the vowel of *met* in American. *Systemic* differences are differences of 'system' or phonemic inventory, whereby a contrast made in one variety of a language is systematically absent in another: a very clear example of this is the absence of any distinction in Scottish English between the vowels of, for example, *Luke* and *look*, though as we just saw, the phonological interpretation of this absent distinction is far from straightforward. *Realisational* differences, by comparison to systemic differences, are differences of phonetic detail that involve no effects on the inventory of phonological contrasts: a good example in English involves the three short front vowels in *bit*, *bet*, *bat*, mentioned above. Finally, *distributional* differences are differences in permitted phonotactic distributions of a given element of the system: the most conspicuous such difference in English is the divide between 'rhotic' and 'non-rhotic' varieties, that is, between those that permit /r/ in syllable codas in words like *car*, *port*, *water* (e.g. General American) and those that do not (e.g. Standard Southern British).

Comparable differences in intonation might be termed *semantic* (differences in the meaning or use of the same tune), *systemic* (differences in the inventory of phonologically distinct tune types, irrespective of semantic differences), *realisational* (differences of detail in the phonetic realisation of the same tune), and *phonotactic* (differences in tune–text association and in the permitted structure of tunes). These will be exemplified as we go along. As just discussed, there are enough conceptual difficulties with defining notions like 'phonologically identical' that no great theoretical significance should be attached to the taxonomy of differences; it is not always easy to classify specific cases, as is indeed often true for the segmental differences on which the taxonomy is based. Nevertheless, there is certainly some heuristic value in being aware that there are different types of differences before we proceed with the case studies in the following subsections.[11]

3.2.3.1 The calling contour

In many European languages, in certain situations, people who are some distance away from a speaker can be called or hailed using a chanted tune on two sustained notes, stepping down from a fairly high level to a somewhat lower level. The extent to which the tune is 'chanted' rather than spoken is both hard

11 The possibility of describing intonational variation in terms explicitly analogous to Wells's was suggested by Francis Nolan in discussion at the third ToBI workshop on prosodic transcription held at Ohio State University in June 1993.

to define and variable; the interval between the two notes is often, but by no means necessarily, three semitones. That much seems to be shared by all the languages that use this 'calling contour', but there are subtle differences from one language to another, which provide nice examples of several of the types of differences just listed.

The basic vocative use of the calling contour varies slightly from language to language. In North American English the stereotypical use is a parent calling for a child to come home, and this use is probably widespread. In German, however, the same tune is also used for calling between adults, usually with the notes less prolonged than in the North American variety, and sometimes only with wordless utterances like [u u] or [hu u]. In our Wellsian taxonomy of differences, this comparison between English and German involves both realisational and semantic differences. In addition, German also uses the calling contour at close range – that is, for purposes other than calling – in a variety of contexts, such as greetings and leave-takings; such close-range uses are much less common in English.[12] For a detailed discussion of the use and realisation of the calling contour in English and German see Gibbon (1976: 274–87).

There are also significant 'phonotactic' differences among European languages with respect to the calling contour. Before we discuss these, however, we must briefly consider the contour's phonological representation. As noted earlier (section 3.1.4), the calling contour was analysed by Ladd (1978) as part of a subsystem of 'stylised' intonation contours, which are systematically related to non-chanted, non-stylised contours in English. As with Pierrehumbert's proposal that English has downstep, the basic idea that English has a set of 'stylised' contours has generally been accepted, but details of the phonological analysis are still a matter of disagreement, especially the nature of the final boundary tone. For purposes of the discussion in this section, distilling what is common to AM treatments of the calling contour, I will ignore the boundary tone issue and transcribe the calling contour as a sequence of an H tone and a downstepped H tone, using the ad hoc notation H . . . !H. I will return to the details when we discuss rising statement contours below.

In English or German, the first H of the calling contour is clearly a starred (accentual) tone, because it must be associated with the nucleus or most prominent lexically stressed syllable of the utterance. Any syllables preceding the

12 Close-range uses seem especially limited in American English. One striking difference between British and North American intonation is the British use of the calling contour on *Bye* at the close of telephone conversations, compared to the normal declarative fall used in North America. The British usage sounds facetious to North American listeners.

nucleus are a low L in a fully chanted version of the contour, though they may be at a fairly indeterminate mid level in a more 'spoken' version. If there is one syllable after the nucleus, the !H goes with it; if there is more than one syllable after the nucleus, the location of the step down is determined by a variety of factors not treated here (but see section 4.1.4 below, and cf. Leben 1976 and Haggo 1987); if the main lexically stressed syllable is utterance-final, it is in effect split up into two syllables, to provide a place (a docking site, in autosegmental terms) for both the H and the !H tone. This is shown in (8):

(8) (a) (English)

H* !H	H* !H	L H* !H	H* !H
Johnny	Jonathan	Rebecca	Jo-ohn

(b) (German)

H* !H	H* !H	L H* !H	H* !H
Lisa	Ursula	Hans-Jürgen	Ja-an

In Hungarian (see Varga 1989) and French, on the other hand, the H ... !H sequence goes on the last two syllables, irrespective of any accentual prominence, though if there is only one syllable, then, as in English or German, its vowel can be broken into two parts. Thus:

(9) (a) (Hungarian)

H* !H	H* !H	L H*!H	H* !H
Anna	Pa-al	Katalin	(or: Katalin)

(b) (French)

H* !H	H* !H	L H* !H
Monique	Ja-acques	Anne-Marie

The difference between the English/German way and the Hungarian/French way of associating the tones in the calling contour can perhaps best be appreciated by comparing the English and French versions on phonetically very similar names such as *Louise* or *Annie*:

(10) (a) (French)

H* !H	H* !H
Louise	Annie

(b) (English)

L H*!H	H* !H
Loui-ise	Annie

French treats these two names identically, because both have two syllables. In English, on the other hand, the tune is associated with the two names in two

different ways, because *Annie* has lexical stress on the first syllable and *Louise* has lexical stress on the second.[13]

A different kind of variability in the calling contour is found in Dutch: as in English or German, the first H is an H*, associated with the main accented syllable; but following the H* it is possible to have two or even more distinct downward steps, each of which seems to reflect the occurrence of a separate !H tone. Thus a Dutch child calling others back to the den in a game of hide-and-seek may call

(11)

H*	!H	!H
uit-	ko-	men

(lit. 'come out'; in citation form the first syllable bears the main stress and the last syllable is reduced). A corresponding call in German (which might be used in hide-and-seek in the same way) would be

(12)

H*	!H
raus-	kommen

with only a single step. The principles according to which second and subsequent steps appear in the Dutch version are related to the overall metrical structure of the utterance and are treated in detail by Gussenhoven (1993).

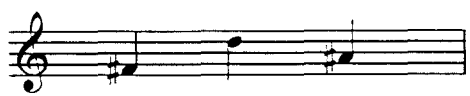
In all the languages discussed here it seems clear that we are dealing with the same tune – again allowing for the conceptual difficulties with the notion of 'the same tune'. It also seems clear that the tune is associated with texts in ways that differ systematically between languages. A comparative study of intonation would obviously be incomplete if it failed to note the striking similarity of the calling contours across the languages of Europe; it may well be – as has often been suggested (e.g. Abe 1962) – that these similarities can be accounted for by functional principles such as the need to sustain high pitch to ensure audibility in calling at a distance. But the comparative study would be equally incomplete if it failed to note the differences of detail between languages. In the present case, the most significant differences of detail are the 'phonotactic' ones discussed in this section – differences that the AM framework makes it straightforward to describe.

13 After a seminar presentation on this general topic, I was once told a supposedly true anecdote of two girls named Louise, one French-speaking and one English-speaking, who grew up together in a bilingual area of Ottawa. The two often played together outdoors, and could always tell which one was being summoned home, by the association of tune and text in the parental call! However, it should be noted that French speakers can also treat *Louise* as one syllable /lwiz/, in which case the calling contour would apply as it does to monosyllabic *Jacques*, and would sound impressionistically much more like the English way of calling the name *Louise*.

3.2.3.2 The French/Italian suspended-fall intonation

French has a characteristic intonation pattern, common in everyday conversation, which may be used in place of ordinary statement intonation to convey an additional nuance. Phonetically the contour can best be described backwards from the final syllable. The final syllable is spoken on a fairly level mid-to-high pitch; the immediately preceding syllable is quite a bit higher; and any preceding syllables are very low. Fónagy, Fónagy, and Sap (1979) indicate the three pitches musically as follows:

(13)



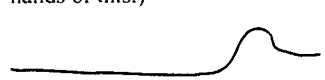
On the basis of the phonetic form, I have called this the 'suspended-fall contour'. Fónagy, Fónagy, and Sap call it the 'declarative triangle' (*triangle assertif*), to distinguish it from the 'interrogative triangle' which they also discuss.

The additional meaning conveyed by the suspended-fall contour is exceedingly difficult to describe in words: it seems to be best characterised as the vocal equivalent of shrugging the shoulders. Its meaning may be similar to that of the English 'surprise–redundancy' contour discussed by Sag and Liberman (1975; see also Bolinger 1982): it seems to convey the idea that the propositional content of the statement is, or ought to be, obvious. Dell (1984: 66) says that the contour 'presents [the sentence] as an incredulous or disapproving reply to someone else's statements' (my translation). Fónagy, Fónagy, and Sap describe the nuance only by saying that it 'recalls the melodic schema of children's taunts (Na-na-na!)' (1979: 9; my translation). Some examples follow.

(14)

(a) Alors il ne viendra plus.

'Well, he won't come any more then.' (possible nuance: I wash my hands of this.)

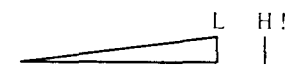


(b) Parce qu'il n'avait plus d'argent.

'Because he didn't have any more money.' (possible nuance: why else?)

Dell (1984) represents this contour transparently as a sequence Low–High–Mid. Assuming a description without a distinctive Mid tone, we might translate Dell's representation into L+H+!H*, that is, a downstepped H accent tone preceded by an L . . . H leading tone sequence. Thus:

(15)



Parce qu'il n'avait plus d'argent.

There is of course no precedent in Pierrehumbert's analysis of English for accents that involve two leading tones, and it may well be possible to show that structurally the L is a left edge tone for the phrase, or belongs structurally to a prenuclear accent; I will not pursue this issue here. However, it is worth mentioning evidence for the H+!H* analysis of the final two tones. There are two main pieces of evidence for this analysis, one from within French and one arising from comparison with Italian.

French makes wide use of the syntactic device of right-dislocation for discourse foregrounding and backgrounding. The following examples, with right-dislocated noun phrases, are quite unremarkable in spoken French:

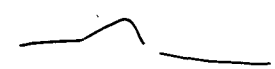
(16) (a) C'est pas joli, cette affaire.
'This business isn't very pretty.' (lit. it is not pretty, this matter)

(b) Elle t'a rien donné, ta mère?
'Didn't your mother give you anything?' (lit. she gave you nothing, your mother)

(c) Mais tu l'as, ton passeport!
'But you've GOT your passport!' (lit. but you have it, your passport)

The intonation of such right-dislocated constituents depends entirely on the intonation of the matrix sentence: if the matrix sentence ends high, the right-dislocated constituent is high, whereas if the matrix sentence ends low, the right-dislocated constituent is low. (Delattre (1966) refers to these two possibilities for right-dislocated constituents as 'echo' (high) and 'parenthesis' (low), and notes that they are in complementary distribution.) Thus:

(17)



(a) C'est pas joli, cette affaire.

(b) Elle t'a rien donné, ta mère?

To this statement we must add the provision that matrix sentences ending with an emphatic or exclamatory high peak (the *intonation d'implication* of

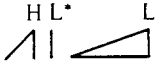
Delattre 1966) are followed by low right-dislocated constituents, because the intonation drops rapidly away from the peak:

(18)

Mais tu l'as, ton passeport!

If we assume that this fall from the peak (which on a single syllable seldom, if ever, reaches the bottom of the range) reflects an underlying intonation pattern H*_L, then the intonation of the right-dislocated constituent in all cases simply involves a copy of the last tone of the matrix sentence. Thus:

(19)



(a) C'est pas joli, cette affaire.

(b) Elle t'a rien donné, ta mère?

(c) Mais tu l'as, ton passeport!

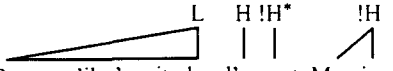
Now consider what happens when the suspended-fall intonation is applied to a sentence containing a right-dislocated constituent. In this case, there is a second downstep: the right-dislocated constituent is lower than the final syllable of the matrix sentence (which in turn is lower than the immediately preceding syllable), but is not as low as the low right-dislocated constituent in (19). Thus:

(20)

Parce qu'il n'avait plus d'argent, Mercier.

This is exactly what we expect if the right-dislocated constituent copies the last tone of the matrix sentence, provided the last tone of the matrix sentence is !H*:

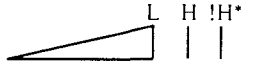
(21)



Parce qu'il n'avait plus d'argent, Mercier.

The second piece of evidence for the H+!H* analysis comes from Italian. Italian has a phonetically similar intonation pattern that is used in similar circumstances with a similar meaning to the French suspended-fall contour just described. In utterances that end on a lexically stressed syllable, the contour is transparently the same as the French one:

(22)

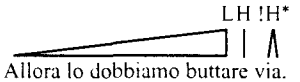


Perché non ne voglio più.

'Because I don't want any more.' (nuance: as you should have noticed)

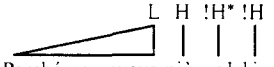
However, the normal position for lexical stress in Italian is penultimate, and earlier stresses are not unusual. When the contour under discussion is applied to sentences with non-final stressed syllables, the result may be a second downstepping level following the stressed syllable, exactly as in French sentences with a right-dislocated constituent. Thus both of the following are possible; the second seems more emphatic (and may be more characteristic of certain regional varieties than others):

(23)



(a) Allora lo dobbiamo buttare via.

'Well, we'll have to throw it away then' (possible nuance: you shouldn't have spoiled it)



(b) Perché non avevo più s o l d i.

'Because I didn't have any more money.'

Up to the stressed syllable, in short, the Italian version of the contour is identical to the French version, that is, a sequence L_H_!H* beginning two syllables before the lexically stressed syllable. But given the Italian stress patterns, the 'full form' of the contour with the extra level may be seen in Italian any time the accented word ends in an unstressed syllable. In French, the full form shows up only in sentences with right-dislocation.[14]

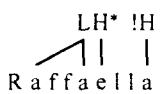
14 The fact that the postnuclear !H tone can be repeated in this way recalls the Dutch version of the calling contour discussed in the previous subsection. I return to the topic of accent-like postnuclear tones in more depth in section 4.1.4.

One final point is worth mentioning. In French, it is not immediately obvious that there is a difference between the calling contour and the suspended-fall contour; that is, an utterance like

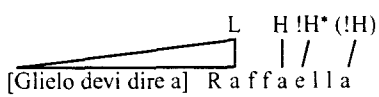
(24)
```
L      H* !H
|      |  |
Anne-Marie
```

appears phonetically ambiguous between the calling contour (see example (9b)) and the suspended-fall contour, conveying a nuance like 'who else would you expect?' The two contours certainly 'feel' different, and speakers are surprised when the similarity is pointed out; but it may nevertheless be that they should be given the same phonological analysis. In Italian, however, the distinction between the two is very clear: the first H tone is the starred tone in the calling contour, while the downstepped !H is the starred tone in the suspended-fall contour. Compare:

(25) (a) *Calling contour*



(b) *Suspended-fall contour*



'You should tell Raffaella' (possible nuance: it's not my problem)

Whether the phonological difference in Italian can be used to motivate a comparable phonological analysis in French (which might therefore explain French speakers' reaction that the two contours are different) is a matter for further investigation.

3.2.3.3 'Rising' statement contours

Our final case study of comparative intonation begins with statement contours in English that are broadly speaking 'rising', but it quickly leads us back to the calling contour. Along the way it presents further examples of the conceptual difficulties involved in cross-language comparison, while also illustrating the value of a cross-language descriptive framework, and the importance of keeping comparative goals in mind when providing AM representations for specific contours in specific languages or language varieties.

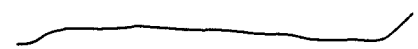
We begin with the 'High Rising Terminal' (HRT) statement contour in English, sometimes known informally as 'uptalk', briefly mentioned in section 1.1.2. This is the phenomenon whereby an intonation pattern that sounds superficially interrogative is used with utterances that are clearly intended as statements. This usage is common in both North America (e.g. Ching 1982) and in Australia/New Zealand (e.g. Britain 1992), but relatively uncommon (though probably spreading) in Britain. It has long attracted a good deal of scholarly attention, especially in Australia and New Zealand (for a useful summary see Warren and Britain 2000), not to mention comment from columnists, cartoonists, and comedians (e.g. Gorman 1993). One common context in which this occurs is in narrative or other monologue, in which speakers often repeatedly use a high-rising contour to invite acknowledgement from the listener; the intonation is a kind of shorthand for 'Do you follow me?' Two other contexts in which the high-rising intonation is often found with statements are transaction-openers and answers to WH-questions. Both can be seen in the following hypothetical exchange:[15]

(26)
```
                H*                  H*HH%
A:  I have an appointment with Dr Macmillan.
B:  What's your name?
      H*     H* HH%
A:  William Jarvis.
```

In both A's opener and his reply to the explicit WH-question, the high-rising intonation means something like 'Are you expecting me? Have you got a record of my appointment?'

In order to discuss this usage sensibly, we need to think clearly about what we want to distinguish it from, and why. First, it is important to note that both of A's utterances remain statements, in the sense that the propositions 'I have an appointment with Dr Macmillan' and '[My name is] William Jarvis' are being asserted, not questioned. The questioning nuance conveyed by the intonation applies to the interaction as a whole, but not to the proposition itself. Compare this with the situation in French, for example, where question intonation on the utterance

(27)



J'ai rendez-vous avec le docteur Desmoulins?

would yield the pragmatically rather odd meaning, 'Do I have an appointment with Dr Desmoulins?' This means that we cannot simply say that American

15 I annotate the HRT contour here as H*_H_H%, but this is actually a matter of uncertainty, for reasons discussed further in section 4.2.2.

or Australian utterances with rising intonation like this are a kind of question; the situation is considerably more subtle than that. On the other hand, we do not simply want to say that they are ordinary statements, either. If we do, we have no explanation for the fact that this usage is noticed and commented on by ordinary language users.

Once again, then, we face the question: what does it mean for two contours to be 'the same' in the system of the language? In the specific case of HRT statements, does the questioning nuance indicate that what we have in these cases is simply question intonation applied to a statement? That is, do A's utterances have the same intonation that he would have used in the same context on a question like

(28) Is this Dr Macmillan's office?

If not, how do they differ? Several studies from Australia and New Zealand (e.g. Fletcher and Harrington 2001; McGregor 2005; Warren 2005) have suggested that the two patterns are not actually phonetically the same, and have documented differences of overall pitch range or alignment that appear to correlate with the difference between HRT statements and real questions. Yet the differences are subtle, and arguably gradient. As we shall see in section 4.2.2, it may be plausible to regard any differences as a matter of paralinguistic variation, and to analyse HRT statements as having the same phonological representation as high-rising question contours.

One reason for treating HRT statements as closely akin to questions emerges more clearly when we consider other superficially 'rising' statement contours in English, in particular the rising or rising–falling intonation typically used in statements in 'Urban North British' (UNB) English (Cruttenden 1994a).[16] Bolinger (1978: 510) simply equated UNB statement rises with American/Antipodean HRT statements, and explained the UNB pattern in terms of the supposed universally interrogative meaning of 'rises'. Specifically, he suggested that UNB rises represent the fossilisation of a conventionalised but formerly meaningful questioning attitude – that the rise had 'a value at one time, now lost'. He based this suggestion on American HRT statements:

16 UNB is the cover term proposed by Cruttenden (1994a) for the varieties of English spoken in Belfast and Glasgow (and Northern Ireland and western Scotland generally), together with the varieties of several major English cities, in particular Birmingham, Liverpool, and Newcastle. These varieties of English share a number of peculiarities, which are sometimes attributed to Irish influence (see Wells 1982, vol. II: 371; Bolinger 1989: 34). This explanation is not implausible, as the three English cities named, together with Glasgow, were major destinations of Irish immigration during the nineteenth century. However, Cruttenden (1994a) provides evidence suggesting that the intonational peculiarities typical of Newcastle were already present in the eighteenth century.

> Many speakers of American English in giving a running account of something will use exactly this kind of terminal rise at the end of practically every sentence – clearly a channel-clearing device that says, in effect, 'Are you listening?', for unless one gives a sign of attention, the monolog comes to a halt. It would not be hard to imagine such a habit becoming a contagion, after which, with interlocutors weary of giving the countersign, the language could be said to have a rising intonation as a mark of clause terminals in general.

Cruttenden, however, shows clearly that the UNB rises are not at all the same as the HRT statements; there are both pragmatic and phonetic differences. Pragmatically, the HRT statements represent a linguistic choice: statement intonation in North American and Antipodean varieties can be either rising or falling in many contexts, and the choice of contour conveys a nuance. This does not appear to be true for the UNB varieties: the rise is not a matter of a statement being pronounced with a question tune, but simply the ordinary way to pronounce a statement.[17] Phonetically, too, the difference is quite striking, in ways that appear to demand distinct AM representations: the American/Antipodean contour keeps rising from the accented syllable to the end of the phrase, whereas in the UNB contour the rise on the accented syllable is usually followed by a distinct fall, sometimes but not often all the way to the bottom of the speaker's range. This can be seen in figure 3.8.

Cruttenden's traditional British analysis of the UNB contour (1986: 139ff.) is as a 'rise–plateau–slump', that is, a rise on the accented syllable, followed by a gradual fall or a steeper fall near the end of the phrase. In AM terms, an obvious phonological analysis would therefore be L+H*_H_L%, or perhaps L*+H_H_L%, or simply (as proposed by Mayo, Aylett, and Ladd 1997 for transcribing Glasgow English) L*_H_L% – that is, a low or low-rising nuclear pitch accent, a high phrase accent, and a low boundary tone. This is the same analysis we gave to Hungarian question intonation in section 2.5, which suggests that UNB statement rises involve a contour type that does not occur in most other

17 It is not even clear that there is a difference between ordinary statement intonation and ordinary question intonation in these varieties of English. Native speakers of Glasgow-area varieties among my students in the early 1990s (in particular, unpublished undergraduate or master's level dissertations by Ann Hastings, Sarah Cole, and Adam Huffman) appear to feel that there is not such a difference. Limited experimental investigation suggests that, in the absence of syntactic cues, the most reliable indicator of whether an utterance is a statement or a question is the height of the H – the higher it is, the more likely the utterance is to be perceived as a question (see further McClure 1980). However, there is considerable variation among the UNB varieties themselves, and it may not be appropriate at all to group them together. For example, there is some reason to think that Birmingham and Liverpool, though they have statement intonation similar to Glasgow, do have a clear distinction between statements and questions, with L*_H_L% a possible analysis of the statement, and L+H*_L_L% a likely analysis of the question (cf. Knowles 1973).

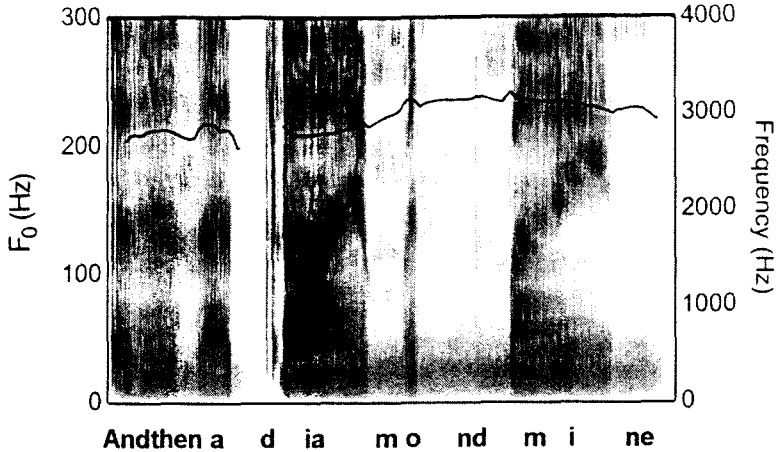


Figure 3.8. *The English phrase* and then a diamond mine, *spoken as the final part of a neutral declarative statement by a Glasgow speaker of Scottish Standard English. The contour is the one called 'rise–plateau–slump' in Cruttenden's British-style analysis, with a low-rising nuclear accent on the first syllable of* diamond, *followed by a relatively high but gradually declining stretch of pitch to the end of the utterance. The utterance is taken from the HCRC Map Task Corpus (Anderson* et al. *1991); the speaker is describing the layout of landmarks on an illustrated map that she can see but her addressee cannot.*

varieties of English. Once again, the AM representation gives us the possibility of getting beyond unfalsifiable claims about intonational universals based on loose descriptions like 'rising contour', and making more focused comparisons based on a more principled phonological analysis.

However, this analysis raises a new problem, and brings us back to the phonological analysis of the calling contour raised in section 3.2.3.1. In both the Beckman-Pierrehumbert 1986 version of Pierrehumbert's analysis of English, and in MAE_ToBI, the sequences L+H*_H_L% and L*+H_H_L% are used to represent variants of the calling contour; and in all versions of Pierrehumbert's analysis the edge tone sequence H_L% is used throughout the tonal inventory to represent the phrase-final sustained level pitch of all 'stylised' contour types. Pierrehumbert's analysis is based on the assumption that after an H phrase accent the only possible contrast is between a rise (H%) and a sustained level, and that the latter can therefore be represented by L%. A more phonetically transparent transcription for the sustained level final pitch would involve a special 'level' boundary tone 0% (Grabe 1998), or simply the absence of any boundary tone whatsoever (Ladd 1983a; Gussenhoven 2005). We therefore

need to consider the apparently incompatible uses of the H_L% edge tone sequence in Pierrehumbert's analysis of stylised intonation and the analysis just proposed for UNB statements.

One possible answer would be that the two analyses are not incompatible. This was suggested by Pierrehumbert in discussion at the 1993 Ohio State ToBI workshop. In this view, what counts is the set of possible contrasts within any given dialect. In American English, so the argument goes, the pitch can either stay level or rise after an H phrase accent, but cannot fall. Therefore, for American English, a two-way opposition L% versus H% expresses the phonological possibilities. If in UNB English there are similarly only two possibilities – either fall versus level or fall versus rise – then L% versus H% will work equally well for UNB as for American. The difference between the two dialects would be one of phonetic realisation rather than of phonological system. That is, H_L% would represent final sustained level in American and final fall in UNB, while H_H% would represent final rise in American and either final level or rise in UNB.[18]

Yet although the theoretical logic appears sound, Pierrehumbert's proposal undermines the value of the AM descriptive framework for comparative discussion of intonation. If we give identical phonological analyses to markedly different contours, it makes cross-language and cross-dialect comparison of the sort I have been exemplifying at best difficult and at worst meaningless. As I noted earlier, the entire comparative enterprise hinges on being able to treat identical AM analyses across dialects and languages as representations of 'the same' tune. Specifically, if we follow Pierrehumbert's suggestion that there are differences in the phonetic realisation of the H_L% in different varieties of English, we would then have to treat the UNB calling contour as phonologically different from the American calling contour; worse, we would have to treat the UNB statement contour and the American stylised low-rise as being phonologically the same. Yet the comparison in section 3.2.3.1 depended on being able to use the same basic analysis of the calling contour in numerous languages of Europe. Unless it can be shown that UNB intonation is wildly deviant in a European context, it would seem preferable to assume that its calling contour is tonally like those in other European languages. More generally, if comparison of the sort presented in this chapter is of any value at all, it implies that we

18 Unfortunately, there are probably three possibilities after H phrase accent in UNB, namely rise, sustain, and fall. I say 'probably', because the distinction between sustained level and rise is perhaps not as clear as in most other varieties of English. One might, for example, argue that the only occurrences of actual rises in UNB represent dialect borrowing from RP, or that the only occurrences of sustained level represent incomplete utterances with no boundary tone. I do not think that these arguments can be maintained, but the situation is certainly not clear-cut.

should not give identical phonological analyses to markedly different contours in mutually intelligible varieties of the same language.

The only way to avoid identifying UNB statements with the American calling contour is to reanalyse one or the other. In my view it is clearly the latter that should be changed. The calling contour (like stylised intonation in the sense of Ladd 1978 more generally) is quite special, and, in a way, marginal in the intonation system as a whole. This marginal status is not captured if we analyse it, as Pierrehumbert does, with a sequence of phrase accent and boundary tone that just happens to remain unused when the other contrasts of the system are taken care of, and that requires specific comment about phonetic realisation. As I suggested above, I believe that the most plausible representation of sustained phrase-final pitch is one without a boundary tone or with a Grabe-style 0% boundary tone, possibly with tone spreading to associate each syllable with a distinct level in the tonal string. Such a representation would leave us free to analyse the UNB calling contour in the same way as in other European languages, and to use the phonetically transparent H_L% sequence for the UNB statement tune without entailing intuitively implausible consequences for the comparison of different varieties of English.

3.2.4 Summary

In section 2.5, using the example of Hungarian question intonation, I illustrated what I see as the rich potential of the AM approach for understanding the nature of intonation cross-linguistically. I said that if we really want to understand what is universal about intonation, we have to start from descriptions of individual languages that are expressed in comparable terms – terms that are precise enough to serve as the basis of testable hypotheses – and I suggested that the AM framework gives us the tools to create such descriptions. In section 3.2, however, I have gone on to suggest that we may throw away the advantages conferred by the AM approach if we insist on too narrow an interpretation of what the approach entails. I have specifically argued that, for comparative and typological purposes, it is not helpful to insist that each language should be analysed in its own terms. It is important to be precise about what we are comparing, but it is also important to recognise that the necessary precision resides not just in devising ways of representing the apparent intonational contrasts of a specific language. Part of the typological task requires us to find proper definitions of fundamental AM concepts like phrase accent and nucleus – and, indeed, of the very notion of intonational contrast. By deciding prematurely that we have solved one language, we may miss the opportunity to make real progress in understanding something about language in general.

4 *Issues in AM pitch phonology*

We have now seen in some detail how the AM theory provides a useful descriptive framework that has been successfully applied to many languages, and how it gives us the potential to make more specific and hence more testable statements about intonational universals. To the extent that we discussed disagreements about the appropriate analysis of specific contour types, the discussion was framed in the context of cross-language comparison: we argued that certain analyses are more appropriate than others in view of the goal of linguistic typology. However, many of the disagreements actually depend on unresolved theoretical issues within intonational phonology itself, questions that are relevant whether we are comparing cross-linguistically or trying to describe a single variety of a single language. These are the topic of this chapter.

4.1 Issues in tonal description

4.1.1 Nuclear and prenuclear accents

We begin with an issue that lies at the boundary between 'pitch' and 'prominence' (the two aspects of intonation sketched in figure 1.1), namely the distinction between 'nuclear' and other kinds of accents. This distinction, in one form or another, has been a part of theorising about intonational structure at least since the beginning of the British tradition. (See Cruttenden 1992b for an interesting discussion of pre-twentieth-century precursors to the nucleus idea.) According to the founding work in the British school (Palmer 1922), an intonation contour is divided into three parts, called head, nucleus, and tail. Only the nucleus is obligatory, so that on a monosyllabic utterance the contour consists of the nucleus alone. In an utterance with more syllables, the nucleus occurs on the most prominent stressed syllable, which is normally also the last stressed syllable. One of a small set of different 'nuclear tones' (four in Palmer's analysis) can be selected at the nucleus. The shape of the tail – the stretch of contour following the nuclear syllable – is largely or entirely dictated by the

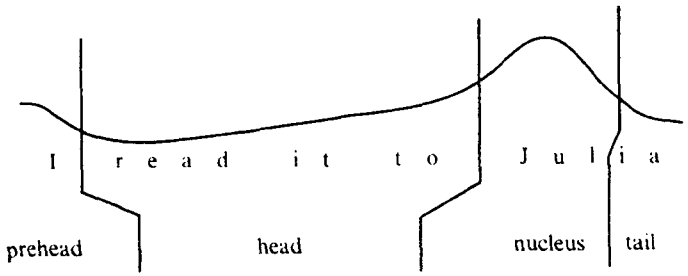


Figure 4.1 *Division of the pitch contour in the typical analysis of the British school.*

choice of nuclear tone. The shape of the head – the stretch of contour preceding the nuclear syllable – is an independent choice, though many in the British tradition (e.g. O'Connor and Arnold 1973) have noted that some combinations of head and nuclear tone are particularly favoured or disfavoured.

With some variation, this basic plan has been part of the British approach to describing intonation ever since. The inventory of contour types for the 'head' has been the subject of considerable disagreement, as has to a lesser extent the inventory of nuclear tone types, but the basic structure has scarcely been altered. Probably the most significant widely agreed change to the structure outlined by Palmer has been the idea that the head begins with a major stressed syllable (see Kingdon 1958), which entails the addition of the 'prehead' – that is, any unstressed syllables preceding the head – to the structure of the contour. Even Crystal (1969) and Halliday (1967a), who provided penetrating theoretical criticisms of many of the assumptions of the British tradition, retained the idea of the distinction between the nucleus ('tonic' in Halliday's terminology) and other parts of the contour. The basic plan of the division of the contour according to the British school is shown in figure 4.1.

Similar ideas are found outside the British tradition as well. For example, Pike (1945) drew a distinction between 'primary contour' and 'precontour' that almost exactly matches the distinction between nucleus (or nucleus-plus-tail) and head (or prehead-plus-head). Both Hockett (1955, 1958) and Trager (1964) explored various versions of the same distinction in connection with the American structuralist theory of intonation. The IPO tradition, though it originally had no such distinction, eventually incorporated a three-way distinction between prefix, root, and suffix, which is comparable in many respects – though not identical – to the British distinction between head, nucleus, and tail ('t Hart, Collier, and Cohen 1990).

As we saw in section 3.1, however, Pierrehumbert's original analysis of English denied any validity to this distinction, and regarded the nuclear accent as nothing more than the last accent of the phrase. Specifically, Pierrehumbert rejected the head/nucleus/tail division: contours are strings of accents, generated by the finite-state grammar in figure 3.1. Pierrehumbert did continue to refer informally to the last accent in the sequence as the 'nuclear' accent. She also acknowledged a substantial correspondence between the British taxonomy of nuclear tone types and what for her are sequences of the last accent and the following phrase accent and boundary tone (see section 3.1.2 and table 3.1). Nevertheless, she denied the notion of nucleus any formal status in the theory, and this view was initially widely accepted by others in the AM tradition.

It is worth noting, however, that Pierrehumbert's argument against the nucleus is in effect mostly an argument against the 'head' of the British tradition, an argument that follows from her argument against global contour shapes. That is, Pierrehumbert denies that the head is an identifiable component of contours primarily because traditional British descriptions of the head are often expressed in terms of global shapes ('scandent head', etc.). As I have argued elsewhere (Ladd 1986), it is important to distinguish arguments against the head as a global shape from arguments against the head as a substring that precedes the obligatory nucleus. It is entirely possible to distinguish the nucleus from other accents – and hence to recognise implicitly the existence of the head – without abandoning Pierrehumbert's basic assumption that intonation contours are strings of pitch accents or tones. There is no deep incompatibility between AM assumptions about pitch and the idea that the nucleus has a special status.

This is because the true basis of the special status of the nucleus is its 'primariness' – the fact that it is the only obligatory pitch accent in the phrase. This is what gives it an important role in signalling pragmatically prominent words and constituents, as we shall see further in chapters 6 and 7. This conception of the nucleus is made more plausible by the revised Beckman-Pierrehumbert conception of the phrase accent as the edge tone for the intermediate phrase (cf. section 3.1.4.1), which means that the nuclear accent is the last accent in the intermediate phrase, not the intonation phrase. We can therefore easily talk of local peaks of relative prominence defined within short strings of accents even in a longer utterance.

The fundamental compatibility of the notion of nuclear accent with other AM assumptions has increasingly been acknowledged in the literature. It is now common to find references to 'nuclear' accents, and, as noted in section 3.1.2, the online ToBI training materials not only explicitly define the nuclear accent as the last accent in the intermediate phrase, but state that it is more prominent

or more salient than prenuclear accents. The British tradition's terminology is thus substantially redefined in accordance with AM assumptions, but the basic descriptive insight that every intermediate phrase has its own peak of prominence is retained. We will return to this issue in chapter 7, when we consider the phonological aspects of phrasing and prominence; for the rest of this section we deal with issues that are more clearly about pitch.

4.1.2 *The identification of tones*

In section 2.3, we saw that one of the central claims of the AM approach is that pitch accents are composed of combinations of H and L tones. Yet some of the issues discussed in chapter 3 revolve around fundamental questions of how tones are to be recognised. By what criteria do we decide that a given pitch accent consists of one or two (or even three) tones? How do we determine that there is or is not a tone at any given point in a string?

In Bruce's thesis (1977), which, as we saw in chapter 1, represents one of the important foundations of the AM approach, tones are identified with *turning points* in the F_0 contour. Local maxima correspond to H tone and local minima to L tone. This is a very 'concrete' conception of tone, and leaves little room for argument about what the tones are in a given contour. It does leave a little room, because local maxima and local minima that are present physically in the contour may be attributed to segmental perturbations of F_0 or other non-intonational causes. Also, Bruce (1977: ch. 5) explicitly discusses the possibility that tonal targets may be undershot in an H ... L ... H sequence, for example, where the targets are very close together. On the whole, however, identifying tones with turning points puts severe limits on the range of possible phonological interpretations of a given contour.

This concrete conception of tones as turning points is implicitly abandoned by Pierrehumbert. In her approach, as we have seen, tones need not always correspond to turning points, and turning points need not always reflect the phonetic realisation of a tone. We saw an example of the latter case in section 3.1.3: the low turning point between two H* accents is not taken to reflect an L tone of any sort, but instead is accounted for by phonetically modelling the H* as a local pitch jump rather than merely a peak, with a 'sagging transition' between the two accents. In the same section we also saw examples of the converse case, where posited tones do not correspond to turning points. For example, the L in an H*+L accent is never realised as a low target or turning point, but serves only to trigger downstep on the next H tone. The L* in an H+L* accent probably represents a target, but never a low turning point: it is an accentual target, lower than the immediately preceding context, but scaled

at the level of a downstepped H tone.[1] Similarly, the L% in Pierrehumbert's analysis of English stylised contours does not represent a final low target, but only the absence of a further rise following the H phrase accent (section 3.1.4.2 and 3.2.3.3).

A priori there is no reason to reject Pierrehumbert's approach, or to prefer Bruce's. It is abundantly clear from the tonal phonology of many African languages that we must allow for the possibility that tones in the underlying phonological string may not be realised phonetically as distinct F_0 targets. The clearest cases, as in Pierrehumbert's analysis of English, involve downstep. In many languages, the L in an H ... L ... H sequence may have no independent realisation as a turning point, but may be manifested only in the lowering of the second H relative to the first. Other cases are less clear, but still definitely involve a more complex relation between tones and targets. For example, Laniran (1992: ch. 3; Laniran and Clements 2003) shows that in Yoruba, the sequences L ... H ... H ... H and H ... L ... L ... L are not realised with abrupt steps from L to the first H, or H to the first L, but rather involve a gradual transition from L to the third H, and from H to the third L, almost as if the intervening syllables are tonally unspecified. There is much we do not understand about the relation between phonological tones and F_0 targets, but it is clear that the simple equation 'tone = turning point' is too restrictive to serve as a universally valid principle of tonal realisation.

At the same time, in dealing with languages without lexical tone, there are reasons not to proceed too enthusiastically to posit tones that are realised in indirect ways. For one thing, by abandoning Bruce's simple equation we throw away a powerful empirical check on our theorising. In languages with lexical tone, we often have independent evidence for postulated tones that do not surface as target points, because we can observe alternations in the shape of individual morphemes and words in different phonological contexts. In languages with only postlexical or intonational uses of tone, we have no such ready way of constraining our descriptions.

A clear illustration of the problems of abandoning the identification of tone and turning point is provided by an early proposal of mine (Ladd 1983a) for a 'delayed peak' feature. I suggested that Bruce's 'alignment' analysis of the Swedish Accent 1/Accent 2 distinction could be extended to describe intonational distinctions in other languages in the context of an AM tonal

1 I refer here to Pierrehumbert's original use of the transcription H+L* (i.e. to what became H+!H* in the ToBI standard), not to the common use of H+L* in transcribing the Romance languages, German, etc: cf. sections 3.1.3 and 3.1.5.

analysis. Specifically, I suggested that the rise–fall or 'scooped' accents in English (see e.g. Vanderslice and Ladefoged 1972: 822; Ladd 1980a: 35 and 110–12) are identical in tonal composition with ordinary falling accents, and differ only in the way the H tone is aligned with the accented syllable. I analysed both tonally as H*+L, but proposed that in the ordinary fall the peak is aligned early in the syllable, whereas in the 'scooped' fall the peak is aligned considerably later, sometimes well into the following unstressed syllable ('delayed peak'). I argued that this analysis expresses the functional similarity between the two accent types (which is reflected in the failure of traditional analyses to distinguish them consistently), yet at the same time provides a phonetically explicit description of the difference. However, the description is actually not very phonetically explicit: it ignores the fact that the scooped contours often involve a slight stepping down to a low turning point on the accented syllable before beginning the rise to the (delayed) peak and subsequent fall. This low turning point, I now acknowledge, is almost certainly to be interpreted as a low accentual target aligned with the accented syllable, namely an L*; that is, the delayed peak contour should be treated as L*+H, followed by a further L target (e.g. a phrase accent). By concentrating on the alignment of the peak and on my arguments for treating all falling accent contours as H*+L, my proposal allowed no way to account phonetically for the low turning point.

A more extensive illustration of the problems with ignoring turning points is provided by the case of the 'sagging transition' between H* targets in Pierrehumbert's analysis of English. In Bruce's original conception, transitions between tonal targets are simply straight-line interpolations. Pierrehumbert's own work has shown the value of this idealisation in describing tonal realisation in Japanese (Pierrehumbert and Beckman 1988: esp. ch. 4). But in her analysis of English, as we saw in section 3.1.3, she proposes that the transition from one H* accent to the next is not a straight interpolation between the peaks, but gradually sags or declines after the first H* until just before the second H*, at which point there is a local rise to the level of the second peak.

This analysis, on its own, has a certain amount to recommend it, but considered in the light of the overall description it makes for problems with the interpretation of turning points, as Pierrehumbert herself acknowledges. For example, in a two-accent sentence like *Their mother's a lawyer*, with identical high peaks on both accents, we have in theory a distinction between H* ... L+H* and H* ... H*. The sequence H* ... L+H* has a smoothly dropping transition from the first peak to a fairly low level immediately before the second peak, while the sequence H* ... H* shows a sagging transition that falls gradually

until just before the second H*, at which point it rises slightly but abruptly. The obvious question raised by this analysis – assuming we can reliably distinguish the two contours – is why the gradual fall and abrupt rise reflect an L target (the leading tone of the L+H* accent) in one case, but no target at all – only a 'sag' – in the case of the H* ... H* sequence. This question was investigated in detail by Ladd and Schepman (2003) and Dilley, Ladd, and Schepman (2005), who showed clearly that in all such contours – including ones transcribed by experienced ToBI transcribers as H* – the beginning of the rise is abrupt and more or less simultaneous with the onset of the accented syllable. (For example, the valley between the given name and the family name in names like *Norman Elson* and *Norma Nelson* was timed differently depending on the syllable membership of the medial /n/.) They also showed that the F_0 level of the beginning of the rise was extremely consistent from context to context, and did not vary with the distance between the two accents as predicted by the sagging transition analysis. Everything in their data, in short, pointed to the conclusion that the beginning of the rise is a tonal target whose timing and F_0 level are carefully controlled.

Pierrehumbert recognised that the sagging transition is an unfortunate feature of her theory, and stated that she 'made a serious attempt to get rid of it by developing an account under which the dip in [such contours] arises from a L tone' (1980: 70). However, given her other analytical decisions, none of the other conceivable sources of L was available. In particular, the most serious obstacle to an intuitively satisfying analysis of the sagging transition was her abstract tonal treatment of downstep. As noted in section 3.1.3, both Pierrehumbert's original analysis of English and the Beckman-Pierrehumbert revision treat downstep as a matter of the phonetic realisation of certain sequences of tones. One of the sequences of tones otherwise suited to describing the sagging transition, namely H*+L ... H*, is pre-empted for use in triggering downstep; the other, namely (L+)H* ... L+H*, would trigger downstep in Pierrehumbert's original analysis, and would run afoul of the definitions of L+H* in the Beckman-Pierrehumbert revision. However, if downstep is treated as an orthogonal modification of pitch range that may be applied independently of specific strings of tones (as in ToBI descriptions generally), these problems disappear, and the possibility of treating the 'sagging transition' as involving an actual L tone becomes more attractive.

In any case, this case shows that decisions about what constitutes a tone are not made in an analytical vacuum. By concentrating only on what I saw as Pierrehumbert's failure to produce an 'insightful taxonomy' of falling accents, my analysis of the 'delayed peak' contours failed to realise the potential of the

AM approach for detailed description of phonetic targets. By giving priority to her analysis of downstep, Pierrehumbert was unable to make sense of the turning point preceding the second H* target in a sequence of two high accents. In both cases, the equation 'tone = turning point' was abandoned because other considerations came first. Either way, the definition of 'tone' requires clarification.

4.1.3 *Phonological organisation of tones*

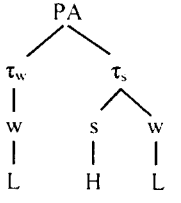
In the case of the sagging transition and in many others, the conflicting considerations in the phonological treatment of phonetic turning points mostly involve *phonological organisation*. One of the central claims of Pierrehumbert's work is that tones can be organised into larger structures – pitch accents, tunes, and so on. At one level of description, the F_0 contour is the realisation of a string of tones, but at another level it is necessary to recognise that the string has internal structure. If we decide that a given phonetic turning point represents a tone, then the very first thing we need to do is to decide how it fits into a larger phonological structure. For example, in order for Pierrehumbert to recognise the low turning point between H* accents as the reflex of a tone, she would have had to find a place for it in a pitch accent type. But the number of possible pitch accent types was extremely limited by her *independent* theoretical assumption that pitch accents never consist of more than two tones. It is because of this assumption that Pierrehumbert's analysis of downstep interfered with recognising the low turning point as a tone, and favoured the 'sagging transition' analysis instead.

If we distinguish carefully between the identification of tones and the analysis of their phonological organisation, we can see more clearly which aspects of an AM analysis rest on fundamental assumptions and which are essentially hypotheses cast within a framework of such assumptions. Consider again the English rising–falling–rising contour that was applied to *Sue* and *a driving instructor* in examples (1) and (2) in chapter 2. It is required by the basic assumptions of the AM approach that there are at least four tones in this contour: an initial L, an H peak, a valley L, and a final H. We might want to argue for the presence of two Ls in the valley – one immediately following the H peak and one immediately preceding the final H; we might also have more to say about the stretch of contour preceding the H peak; but four tones are a minimum. Given those four tones, however, it is a specific descriptive hypothesis that they are organised into an L+H* pitch accent, an L phrase accent, and an H% boundary tone. Other analyses are possible.

For example, a classical British-school analysis of this contour as applied to *a driving instructor* might be as a fall–rise 'nuclear tone' and a low 'prehead'. In effect, the H peak, valley L, and final H are grouped together into one unit, while the initial L is a reflection of a separate unit. There is actually a considerable lack of clarity within the British school about how nuclear tone contours spread to cover more than one syllable (in this case, the nuclear syllable *driv-* and the syllables *-ing instructor* of the 'tail'), but there is no disagreement that the fall–rise constitutes a single unit in the system, distinct from the prehead. We might therefore translate this British analysis into AM terms as involving an L prehead and an H*LH nuclear accent – which is essentially equivalent to the AM phonological proposals of Gussenhoven (1984). Both analyses involve the same L ... H ... L ... H sequence of tones, but they structure the sequence in quite different ways.

The distinction between recognising the existence of a tone and hypothesising about its place in the phonological structure of the tune is explicitly discussed and clearly illustrated in the work of Martine Grice (Grice 1995a, 1995b). Grice shows that the nuclear or most prominent syllable of a short question in Palermo Italian is always marked by a high peak followed by a fall; she also shows that the peak (unlike peak + fall nuclear tunes in English, or statement tunes in Palermo Italian) must be preceded by a valley or a low level stretch of contour. We are therefore dealing with a tonal sequence L ... H ... L. As Grice points out, in the British tradition, the first L would inevitably be assigned to the 'head' (i.e. the stretch of contour preceding the nucleus), and the nucleus would be analysed as a fall or a rise–fall. She argues instead that the L is an essential part of the question intonation nucleus, and posits the following structure for the nuclear accent:

(1)



In the terms used here, the issue between Grice and the British tradition is not about the existence of the L tone before the nuclear H peak, but about its phonological status: does it belong to the nuclear contour, or to the preceding prenuclear stretch?

The issue is not limited to the constituency of the L tone, however. In proposing a structure like (1), Grice also breaks new theoretical ground within the AM approach, which she justifies at some length. That is, the issue between Grice and a standard Pierrehumbert-style analysis is not what the tones are, nor even whether the leading L is part of the nucleus, but how many tones can be organised into a pitch accent. Pierrehumbert proposes a flat structure for pitch accents, with one obligatory starred tone and at most one unstarred tone. If we adhere to this structure, the Palermo question nucleus will have to consist of an L+H* pitch accent followed by an L phrase accent. Grice proposes a more elaborate structure for pitch accents, which can accommodate three or even four tones in a single accent; in her analysis, all three tones of the Palermo question nucleus are part of the same pitch accent.[2]

Related remarks can be made about Gussenhoven's notion of 'tone linking', which he developed in his early work on both Dutch and English (e.g. Gussenhoven 1984), and which still forms part of the ToDI system (Gussenhoven 2005). Gussenhoven proposes that there are various possible realisations of the component tones in any given sequence of accents, depending on phrasing and speech rate. Specifically, he suggests that it is possible to move or delete certain individual tones to link pitch accents more closely. In *partial* linking, the trailing tone of one pitch accent is reassociated as a leading tone to the following accent; in *complete* linking, the trailing tone is deleted. Gussenhoven thus proposes to relate the following three sequences as shown on his example sentence *Toronto is the capital of Ontario*:

(2)

		H*L H*L
(a)	basic form:	Toronto is the capital of Ontario
		H* L H*L
(b)	partial linking:	Toronto is the capital of Ontario
		H* H*L
(c)	complete linking:	Toronto is the capital of Ontario

In the context of the present discussion, we can see that Gussenhoven's claim involves two separate aspects. First, he identifies three contours that differ with respect to the phonetic location of a particular low elbow or turning point. Second, he proposes that the three contours are systematically related in the phonology – that they are in some sense different versions of the same thing. In an orthodox Pierrehumbert analysis, there would be no major disagreements

2 In subsequent work (e.g. Grice *et al.* 2005), Grice has modified her analysis of Palermo Italian to bring it more into line with other ToBI labelling systems, but that does not affect the point being made here; nor does it detract from the interest of her earlier theoretical discussion.

over the surface tonal analyses of the three types. Presumably (2a) would be H*_L ... H*_L, implying that it consists of two intermediate phrases; the other two would presumably be H* ... L+H*_L (2b) and H* ... H*_L (2c), that is, in both cases only a single intermediate phrase. However, no phonological relationships of the sort posited by Gussenhoven would be assumed; the contours are simply different. Gussenhoven's analysis of (2a) and (2b) focuses our attention on the L tone between the two H*s: what kind of tone is it, and where does it come from? For Pierrehumbert, it has two completely unrelated sources, and represents two underlyingly different phenomena: in one case it is the phrase accent of a separate intermediate phrase *Toronto*; in the other case it is the leading tone of the accent on *Ontario*. Gussenhoven, by relating the two contours in the phonology, suggests that in both cases we are dealing with the same thing: it is underlyingly associated in some way with the first accent but can be moved around.

A broadly similar analysis of comparable data in German is discussed at length by Uhmann (1988, 1991). Uhmann's approach, like Gussenhoven's, assumes that the L belongs underlyingly to the first accent. She suggests that this L can spread rightward to a greater or lesser extent, with various specified consequences for phonetic realisation; for a case like (2a) Uhmann would have the L tone associated only with the immediately post-stress syllable in *Toronto*, whereas in a case like (2b) the L would spread to all the intervening unaccented syllables. She also explicitly allows the possibility of intermediate cases, which she relates to different possible focus structures. Again, this suggests that some of what Uhmann discusses might be handled by Pierrehumbert in terms of the location of an intermediate phrase boundary, with the L as a phrase accent.

On any reading of AM principles, there is clearly some kind of low tone in between the two H* accents in (2a) and (2b). Pierrehumbert's analysis of English offers an unambiguous way to transcribe the low tone in both contours, making a clear distinction between phrase accents on the one hand, and leading or trailing tones on the other. However, it provides no way to treat the contours as closely related. Gussenhoven's or Uhmann's analyses rather loosely exploit various kinds of phonological rules in ways that may not stand up to closer inspection, but they deserve attention nevertheless. They question the sharp descriptive distinctions drawn by Pierrehumbert, and they suggest that phrase accents and trailing tones might be, in some way, different manifestations of the same thing. This idea is entirely heterodox from the point of view of Pierrehumbert's description of English intonation, but from a slightly broader perspective unquestionably falls within the general scope of the AM approach.

4.1.4 The phrase accent

This brings us, finally, to the issue of the phrase accent. Recall (section 3.1) that falling nuclear accents are analysed in Pierrehumbert's description of English as a sequence of a high (H* or L+H*) accent, an L phrase accent, and an L% final boundary tone. If we accept the redefinition of the nuclear accent as the final accent of the intermediate phrase, we will concentrate only on the combination of pitch accent and phrase accent, and treat the falling nuclear accent as a combination of H* or L+H* pitch accent and L phrase accent. Either way, analysing nuclear accents as a sequence or combination of an accentual tone and a separate edge tone (the phrase accent) has been controversial from the very beginning. A number of authors (including myself in earlier papers) have argued or simply assumed that the most obvious representation of a falling nuclear accent involves an H*+L pitch accent (e.g. Ladd 1983a; Gussenhoven 1984; Féry 1993; Frota 2002), and that there is therefore no motivation for positing a phrase accent in most European languages at all. I now acknowledge that there is plenty of evidence for something like the phrase accent; that is, there is evidence for the idea that nuclear accents consist of two distinct phonological elements – an accentual element and a peripheral element or edge tone.

One important piece of evidence is the fact, repeatedly demonstrated in recent work, that nuclear accent peaks in many languages are aligned earlier than prenuclear peaks. This was shown informally for English as early as Steele 1986, and has more recently been shown in larger phonetic studies, for Spanish by Nibert 2000; for Greek by Arvaniti and Baltazani 2005; for Dutch by Schepman, Lickley, and Ladd 2006; and for both RP and Scottish English by Ladd *et al.* forthcoming. An obvious explanation for this finding, if we acknowledge the existence of phrase accents, would be to say that the nuclear accent peak must be aligned earlier to make room for the upcoming phrase accent. We know that such 'tonal crowding' affects the realisation of intonation contours in different prosodic contexts (e.g. Arvaniti, Ladd, and Mennen 2006; Prieto and Torreira 2007), and the presence of a following phrase accent is certainly one plausible source of tonal crowding. This explanation was first discussed, as far as I know, by Hualde 2002.[3]

Another consideration is the one just discussed, namely Gussenhoven's and Uhmann's suggestions that the valley between two high accent peaks could

3 This is not to say that Hualde's explanation is universally accepted. Various authors (e.g. Face 2002 for Spanish) have proposed to represent the different alignment of nuclear and prenuclear accents directly in various ways.

be considered a semi-independent trailing tone that is underlyingly part of the preceding accent, regardless of the variation in detail in the way it is realised. One might justifiably regard 'semi-independent trailing tone that is underlyingly part of the preceding accent' as a useful definition of 'phrase accent'. If we look past the specific differences of detail, and accept the idea that phrase accents in some slightly broader sense may show variable realisation (tone linking, etc.), then it may be quite straightforward to integrate Gussenhoven's and Uhmann's ideas into a broader understanding of the difference between nuclear and prenuclear accents. Some useful discussion of this issue can be found in Frota 2002.

The most important strand of evidence in favour of the phrase accent analysis is the existence of what can best be referred to as postnuclear accents. Like many ideas about intonation, this has antecedents in Halliday's work. In his analysis of English (1967a, 1970), Halliday posited two 'compound tones': a fall followed by a low rise (Halliday's Tone 13) and a rise–fall followed by a low rise (Halliday's Tone 53). According to Halliday, these compound tones have two tonic (i.e. nuclear) accents, with the second subordinate to the first. In Halliday's words (1970: 43):

> Tone groups with ... Tones 13 and 53 have the same general pattern of information as other tone groups, with the tonic marking new information. The difference is merely that there are two places where the speaker has decided to focus the information in the information unit, instead of one.
>
> The two tonics are unequal in value. The first tonic, with Tone 1 or Tone 5, is the 'major' one; the second, with Tone 3, is the 'minor' one. The major tonic carries the principal new information in the tone group. The second tonic expresses information which is in some way secondary or subsidiary to it.

The principal basis for the idea of 'postnuclear accent' or 'compound tone' in English is the observation that certain tonal events after the nuclear accent are related to stress or prominence in some way. In short utterances where the nuclear syllable is followed by at most a few unstressed syllables, it is not grossly inaccurate to treat falling–rising or rising–falling–rising tunes as involving a single continuous movement from the nucleus through the 'tail' to the end of the utterance, which is what is implied by the British tradition's 'fall–rise (or rise–fall–rise) nuclear tone'. But as we saw with the examples of *Sue* and *a driving instructor* (section 2.1), when the number of postnuclear syllables increases, it becomes increasingly obvious that the contour consists of a nuclear falling or rising–falling accent and a rising boundary movement. Actually, as the number of syllables increases even more, it becomes increasingly plausible

to posit an L accent on the most prominent postnuclear syllable. Consider the following example:

(3) 

I thought she was dancing tonight.

With the intonation intended, this has the most prominent stress on *thought*, and the implication is 'I thought – despite what someone else may have said – that she was dancing tonight, and it turns out I was right.' That is, there is a (rising–)falling nuclear accent on *thought*, and a final boundary rise. Yet impressionistically there appears to be an L accent on *dancing* as well. In much traditional British work, for example, such a sentence might be transcribed *I `thought she was ,dancing tonight*, that is, with a 'high fall' on *thought* and a 'low rise' on *dancing*. An AM translation of this transcription might be

(4) H* L* L H%
I thought she was dancing tonight.

The problem with such a transcription, if we define the 'nuclear' accent merely as the last in the intermediate phrase, is that it fails to indicate that the accent on *dancing* is somehow subordinate to the one on *thought*. This is what is achieved by Halliday's notion of 'compound tone'.

Some notion of postnuclear accent also seems to be useful for describing the calling contour, which, as we saw in section 3.2.3.1, involves an H* accent tone followed by a downstepped H tone. It has been noted by Leben (1976) and Haggo (1987) that the location of the step down to the !H is sensitive to relative prominence; specifically, the !H 'prefers' to occur on the most prominent postnuclear syllable. For example:

(5) H* !H H* !H H* !H
(a) Lunch is ready! (cf. L u - u n c h! and Lunchtime!)
H* !H H* !H
(b) Jonathan's turn! (cf. Jonathan!)
H* !H H* !H H* !H
(c) Taxi's waiting! (cf. T a x i ! or Taxi's here!)

The effect of postnuclear prominence in the Dutch calling contour is even clearer: as we saw in section 3.2.3.1, there can be more than one postnuclear !H in the Dutch calling contour, and Gussenhoven (1993) showed very clearly that the steps occur at metrically strong syllables. This behaviour is consistent

with the idea that the postnuclear !H in the calling contour in English or Dutch is in some sense an *accent*, like Halliday's 'second tonic'.

The phenomenon of postnuclear accents – and their connection to the phrase accent issue – was discussed extensively by Grice, Ladd, and Arvaniti 2000, largely on the basis of what they called the 'Eastern European Question Tune', or EEQT. This is the contour referred to in section 2.5 as 'Hungarian question intonation'; languages that use the EEQT include Hungarian, Greek, Romanian, and Serbo-Croatian, but apparently not Bulgarian or Czech, for example. As we saw in section 2.5, this contour consists of an L* (or L*+H) nuclear pitch accent and an H_L% edge tone sequence. Given the general assumptions of the AM approach, it is clear in Hungarian that the H_L% sequence following the main accent involves only edge tones, and despite its acoustic salience is not a pitch accent. There are two reasons for saying this: first, the H_L% sequence does not make the word on which it occurs focused or prominent; second, and more importantly, the H_L% sequence always occurs on the last one or two syllables and is completely indifferent to the location of lexical stress. However, in other languages that use the EEQT the status of the H_L% sequence is problematical: specifically, the line between edge tones and pitch accents is blurred. Grice, Ladd, and Arvaniti (2000) illustrated this point mainly with data from Greek.

In Greek, if there is enough lexical material following the main accent, then the H of the H_L% edge tone sequence will associate with the lexically stressed syllable of a following word. This is illustrated in the following pair of examples (lexically stressed syllables indicated by acute accent, as in standard Greek orthography).

(6) L* HL%
(a) x o r é v i ? 'Is she dancing?' (lit. s/he-dances)
L* H L%
(b) x o r é v i a p ó p s e ? 'Is she dancing tonight?' (lit. s/he-dances tonight)

The H_L% that clearly serves as an edge tone sequence in (6a) exhibits some sort of link with the stressed syllable of *apópse* in (6b). More or less by definition, a tone that seeks to associate with a lexically stressed syllable is a *pitch accent*; yet, like Halliday's 'second tonic', and like the downstepped H in the calling contour, this 'accent' remains in some important sense subordinate to the L* accent, which must be regarded as the nucleus. In (6b), for example, the focus of the question remains broad: the force of the utterances is not 'Is it tonight that she's dancing?' but simply a neutral question about what is

happening. If we want to focus on 'tonight', we must shift the nucleus onto *apópse* and put a prenuclear accent on *xorévi:*

(7) H* L* HL%
x o r é v i a p ó p s e ? 'Is she dancing *tonight*?'

This suggests that the appropriate analysis of these data is as follows. In many Eastern European languages yes–no question intonation consists of an L* or L*+H accent associated with the stressed syllable of the focused word, followed by an H_L% sequence. In some languages, such as Hungarian, the H_L% is invariably associated with the edge of the phrase. In others, such as Romanian and Greek, the association may be as in Hungarian, as in (6a) or (7). In addition, however, the H of the H_L% may be associated *with the stressed syllable of a word that follows the focused word*, as in (6b). If we define 'pitch accent' as 'element of pitch contour associated with prominent syllable', then in the latter case the H 'phrase accent' can only be regarded as a pitch accent. Yet in all cases, the L* remains the main accent, in the sense that it is the one that signals the focus of the utterance. When the H pitch accent occurs, it is thus in some sense postnuclear.

Similar observations can be made about the location of the postnuclear L in falling–rising tunes in West Germanic languages, as in example (4). It can be shown that the alignment of the low postnuclear turning point or points in a falling–rising contour is influenced by the location of lexically stressed syllables following the nucleus: this was discussed informally on the basis of German data by Grice, Ladd, and Arvaniti (2000), and investigated systematically for Dutch by Lickley, Schepman, and Ladd (2005). Lickley *et al.* showed that falling–rising question contours on place names with two postnuclear syllables differed systematically in detail according to whether the postnuclear syllables were both unstressed with reduced vowels (e.g. *Tongeren*), or whether one of the syllables had a full vowel and hence arguably 'secondary stress' (e.g. *Eindhoven, Hengelo*). The differences were consistent with the idea that the L phrase accent seeks to associate with the secondary-stressed postnuclear syllable.

Finally, and perhaps most importantly, we should note that Halliday's notion of compound tones, though it clearly foreshadows the idea of postnuclear accent, also misses an important point, because it makes no attempt to relate the two compound tones (his Tones 13 and 53) to the ordinary English 'fall–rise' (Halliday's Tone 4). As I suggested above, there is a kind of continuum from short utterances (like *Sue!?*) where the unitary 'fall–rise' analysis seems appropriate, through longer utterances (like *A driving instructor!?*) where the separateness of the main pitch accent and the final rise is apparent, to lexically and syntactically richer utterances in which the separate final rise seems to signal some prominence on a specific postnuclear word (like *I thought she was dancing tonight!*). In the view advanced here, the postnuclear low valley in all these contours is in some sense the same, and, *pace* Halliday, there is no fundamental phonological difference between them. In all cases we are dealing with an L 'phrase accent', which may manifest itself as the beginning of the final rise in a shorter contour, or as a postnuclear accent in a longer one. As with the H tone in the Eastern European Question Tune, that is, this postnuclear L attaches to a lexically stressed syllable if one is available, but otherwise merely forms the valley between the fall and the rise. We return to discuss the connection between phrase accents and postnuclear pitch accents again in section 8.1.

4.2 Intonational meaning in AM description

4.2.1 Semantic compositionality and phonological analysis

While the emphasis of this chapter, and of the book as a whole, is on phonetic and phonological issues, it should be noted that Pierrehumbert's analysis of English intonation has led to attempts to explain the 'meaning' of intonational tunes, and that such explanations have been used as evidence in arguing for one phonological analysis over another. This line of research includes the work of both Mark Steedman and of Julia Hirschberg and her colleagues, which is anchored firmly in Pierrehumbert's analysis of English, and the work of Carlos Gussenhoven and myself, which is essentially a critique of that analysis. While this work has many implications for the study of intonational meaning in general, the most important point for our purposes here is that some of these researchers are taking recourse to facts about intonational *meaning* as a way of shedding light on issues of intonational *form*. That is, Hirschberg and Pierrehumbert have used analyses of the meaning of, for instance, the H phrase accent, in part as a way of arguing that the H phrase accent exists; while Gussenhoven and I have used analyses of the meaning of, for instance, downstep, to argue that a downstep feature (or some other direct phonological representation of relative pitch level of tones) exists.

I think it is fair to say that this line of work begins with the critique of Pierrehumbert's intonational phonology presented in my own paper on 'peak features' (Ladd 1983a). Essentially, the 'peak features' were intended as a kind of cross-classifying device for accent types: accents, in addition to being high or

low, can be downstepped or non-downstepped, delayed or non-delayed, raised or non-raised, and so on. I presented this system as an exercise in 'insightful taxonomy', arguing that such features make it possible to combine 'phonetic specification and linguistic generalization'. For example, as discussed in section 3.2.3.1, I suggested that the impressionistic similarities among downstepping contours are directly expressed by the use of a downstep feature applying to downstepped accents. In Pierrehumbert's analysis, by contrast, downstep is merely the phonetic consequence of the occurrence of a 'triggering' sequence of tones, and downstepping contours share no single phonological property. I argued that the phonological taxonomy underlying Pierrehumbert's tonal descriptions was inadequate, because it represented impressionistically similar contours very differently, and impressionistically very different contours very similarly (cf. the commentary on table 3.1).

In Ladd 1983a, I was not very explicit about the theoretical status of the features I proposed, appealing only to vague notions like 'linguistic generalization' and 'impressionistic similarity'. However, the idea of cross-classifying features in intonational phonology is extensively developed in Gussenhoven's work on English, and the relation of the cross-classifying analysis to a theory of intonational meaning is made explicit (Gussenhoven 1984). Gussenhoven proposes that English has three basic 'tones' (roughly equivalent to accent types in Pierrehumbert's analysis), namely H*L or fall, L*H or rise, and H*LH or fall–rise. These 'tones' are subject to various 'modifications', such as stylisation (see section 3.2.3.1) and delay (see section 2.2.3).[4] Each of the categories of this phonological taxonomy – both the tones and the modifications – is then assigned a basic meaning, on the basis of which specific nuances and specific functions of intonational distinctions can be described. For example, Gussenhoven states that each of the three basic accent types has a different basic discourse function: the fall is used to introduce an entity into the 'background' or shared knowledge of the interlocutors (hence its use as basic statement intonation); the rise is used to be non-committal about whether a mentioned entity is part of the background (hence its use as basic question intonation); and the fall–rise is used to 'select' an entity from the background. Similar descriptions are provided for the meaning of each of the proposed 'modifications',

4 In subsequent work on both English (2004: ch. 15) and Dutch (2005), Gussenhoven reanalyses many of the 'modifications' of the 1984 description in terms of more elaborate sequences of tones, in particular combinations of pitch accents and edge tones (e.g. the difference between plain and stylised intonation is treated as involving the difference between presence and absence of a following boundary tone, rather than as a stylised modification of a basic plain 'tone'). To keep the exposition clear here, I present everything in terms of his original analysis.

which interact with the meanings of the basic accent types: for example, peak delay is said to signal that the utterance is in some way very significant or non-routine. While meanings such as these are obviously extremely general, Gussenhoven provides extensive discussion of how they might (given various reasonable assumptions about the nature of pragmatic inference) give rise to specific nuances in specific contexts. Furthermore, Gussenhoven (1984: ch. 7; see also Gussenhoven and Rietveld 1991) backs up his analyses with the results of perceptual experiments, in which speakers of English were asked to judge the similarity between contours on otherwise identical utterances. With some exceptions, these experiments show that contours which Gussenhoven's analysis represents as similar are perceived as similar by listeners.

In a related development, Ward and Hirschberg (1985) provide an account of the semantics and pragmatics of the English rise–fall–rise nuclear contour (Pierrehumbert's L*+H_L_H%), as in examples like

(8) L*+H L H%
(a) He's a good b a d m i n t o n p l a y e r. (Their figure 1.)
(b) A: Did you feed the animals?
(H*) L*+H L H%
B: I f e d t h e c a t. (Their example 13, from Ladd 1980a.)

Ward and Hirschberg assume a single basic meaning for the contour, which can be paraphrased as 'uncertainty'. They show how listeners arrive at specific nuances in specific contexts by principles of pragmatic inference. They develop this general approach in subsequent work (e.g. Hirschberg and Ward 1992, 1995). Like Gussenhoven, then, Ward and Hirschberg are adherents of what we characterised in section 1.4.3 as the 'Linguist's Theory of Intonational Meaning'. The most important difference between Ward and Hirschberg's work and Gussenhoven's is the former's serious attempt to express their analysis in terms familiar to practitioners of formal semantics.

By itself, Ward and Hirschberg 1985 is not intended as a defence of Pierrehumbert's analysis of English. It uses the Pierrehumbert notation L*+H_L_H% to refer to the contour in question, but does not attempt to extend the analysis to other elements of Pierrehumbert's system. In a sense, by its very concern with a whole contour rather than a single pitch accent, it tends to undermine Pierrehumbert's insistence on the independence of the individual pitch accents and edge tones. However, subsequent work by Pierrehumbert and Hirschberg (1990) combines Ward and Hirschberg's standards of formal explicitness with a concern to describe the basic meanings of the elements of Pierrehumbert's system in particular – the six pitch accents, the two phrase accents, and the

two boundary tones. Similar claims are made in Mark Steedman's work on intonational meaning (1991, 2000; Prevost and Steedman 1994), though Steedman's concern is not with choosing between competing analyses of English intonational phonology. In both cases the choice of pitch accent is seen as conveying 'information about the status of the individual discourse referents, modifiers, predicates, and relationships specified by the lexical items with which the accents are associated' (Pierrehumbert and Hirschberg 1990: 286); phrase accents and boundary tones are similarly determined. Ultimately, this entails that, for example, Ward and Hirschberg's analysis of the meaning of the L*+H_L_H% contour must be derivable from the meaning of the L*+H accent, the L phrase accent, and the H% boundary tone.

Because the 'Hirschberg–Pierrehumbert theory of intonational meaning' has been seen in some quarters as a major innovation made possible by the AM approach (e.g. Hobbs 1990), it is worth emphasising that the principal novelty of Pierrehumbert and Hirschberg's work is the attempt to provide intonational meanings *for the elements of the Beckman–Pierrehumbert analysis of English intonation*. In other respects it is based, like the other work we have just sketched, on the Linguist's Theory of Intonational Meaning. Pierrehumbert and Hirschberg abundantly acknowledge their indebtedness to Gussenhoven, Bolinger, and others; their intended innovation is the claim that intonational meaning is compositional, and specifically that the compositionality can be best understood in terms of the Beckman–Pierrehumbert analysis. In a sense, their paper can be regarded as a response to the Ladd–Gussenhoven critique: Pierrehumbert and Hirschberg (1990: 282f.) accept the idea that similarity of meaning should be reflected by similarity in phonological representation, and they seek to show that the elements of Pierrehumbert's analysis do indeed share meaning in common. That is, they accept the Ladd–Gussenhoven premise, but reject the conclusion that Pierrehumbert's analysis needs revision.

Despite the different opinions here, and despite the fairly clear results of Gussenhoven and Rietveld (1991) suggesting that the Gussenhoven analysis gives a better account of listeners' similarity judgements than does the Pierrehumbert analysis, there has been very little real debate on this issue. I think this is primarily because we know too little about pragmatic inference for the debate to be conclusive. On their own terms, analyses like Gussenhoven's, Steedman's, or Hirschberg and Pierrehumbert's can all be evaluated as reasonably plausible – or reasonably implausible, depending on who is doing the evaluating. However, there is no theoretical framework within which we can undertake a comparative evaluation that would command general agreement.

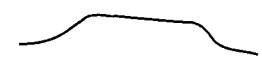
For the present, proposals about intonational meaning are not a reliable source of evidence on intonational phonology.

4.2.2 *Gradient variability and phonological analysis*

One area where meaning seems an especially problematical guide to phonological analysis involves what is often called gradience. As we saw in section 1.4, there are many cases in which an intonational distinction that one analyst regards as involving two ends of a continuum (i.e. based on gradient paralinguistic variation within a single category) is treated by another analyst as involving two distinct categories (i.e. based on a linguistic distinction). In the specific case exemplified in section 1.4, there is uncertainty whether to treat falling nuclear accents as a single category that can be realised with different degrees of emphasis, or to posit two categorically distinct accent types, such as low fall and high fall. The recent AM literature has given rise to several new disagreements of this sort, and it seems clear that so far researchers in the AM tradition have been no more successful at getting beyond what I called 'paralinguistic stalemates' than earlier scholars. There is some progress, in that we now have somewhat clearer *phonetic* evidence of what is gradient and what is categorical (cf. e.g. Ladd and Morton 1997), but there is still something missing in our understanding of the way intonation relates sound and meaning.

One typical – and well-studied – case involves the expression of contrastiveness on sentence-initial topics in German, where there is often said to be a distinction between two types of initial rising accent when the 'hat pattern'[5] (see section 1.2.1) is applied to sentences like the following:

(9)



(a) Am Montag hat's geregnet 'On Monday it rained' (lit. On Monday has it rained)

(b) Am Montag hat's geregnet 'It rained on *Monday*'

In (9a), the accent on *Montag* reaches a pitch peak during the vowel of the first syllable and is often transcribed H*; the adverbial phrase *Am Montag* is merely established as background. In (9b), the first syllable of *Montag* often sounds

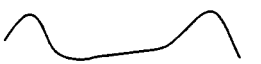
5 In the literature on German this pattern is often referred to as the 'bridge accent' (*Brückenakzent*), a term apparently introduced by Wunderlich 1988. Making due allowance for the issues discussed in section 3.2, it seems accurate to treat the Dutch hat pattern (section 1.2.1) and the German bridge accent as the same thing.

low, and the accentual rise seems to reach its pitch peak in the second syllable, so that the transcriptions L+H* or L*+H seem appropriate; this intonation pattern strongly suggests a contrast with some other day when the weather did something else ('but on Sunday it was sunny'). Discussions of this distinction include Féry 1993; Büring 1997, 1999; and Jacobs 1997; for the transcriptions see the GToBI website (Grice, Baumann, and Benzmüller 2007).

However, experimental investigation of this distinction by Braun (2006) provided no evidence for a clear dichotomy between the two patterns. Her study was based on the speech of nine speakers reading a long series of short paragraphs in which sentences not unlike (9) had been embedded in both contrastive and non-contrastive contexts; speakers were unaware that they were reading identical sentences in different contexts. The results clearly showed three things. First, detailed acoustic analysis of the speakers' production showed significant differences on several phonetic dimensions between accents uttered in a 'contrastive' context and those uttered in a non-contrastive context. For example, contrastive accents had, on average, pitch peaks that were both later (relative to the segmental string) and higher than non-contrastive accents, and the duration of the accented words was longer in the contrastive than in the non-contrastive context. However, these were statistically significant differences between means taking all the accents as a group: there are not two clear clusters of phonetic properties that are unambiguously characteristic of contrastive and non-contrastive accents. Second, in a subsequent perception task based on stimuli with intonation contours resynthesised to reflect the results of the production experiment, German listeners had only limited success in identifying whether accents were intended for a contrastive or a non-contrastive context. This appears to confirm that while there may be some interpretable phonetic differences in the way such sentences are uttered, the differences do not involve distinct phonological categories mapped onto distinct pragmatic interpretations. Third, eight experienced ToBI transcribers gave extremely variable transcriptions of the rising accents in the spoken utterances. There was no consistent representation for either of the two contexts, and little agreement between transcribers; four of the eight transcribers did not usually give different transcriptions to the contrastive and non-contrastive versions. This makes it difficult to maintain an analysis in which two clear pragmatic categories (non-contrastive vs contrastive topic) are associated with two clear phonological categories (H* vs L+H* or L*+H), but at the same time suggests that we cannot simply dismiss the claim that the pragmatic distinction is signalled intonationally.

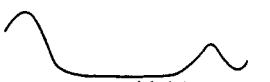
A similar case involves what Liberman and Pierrehumbert (1984) called the 'background–answer' and 'answer–background' contours discussed in section 2.3.3. Example (19) from chapter 2 is repeated here for convenience:

(10) (a) '*Background–answer (BA) contour*'
Question: What about Anna? Who did she come with?



Answer: Anna came with Manny.

(b) '*Answer–background (AB) contour*'
Question: What about Manny? Who came with him?



Answer: Anna came with Manny.

The pragmatic distinction in this case is clear: one of the accented referring expressions is intended as new (Liberman and Pierrehumbert's 'Answer') and the other as given (Liberman and Pierrehumbert's 'Background'). Although there is a clear phonetic indication of this distinction in the Answer/Background order (10b) based on the relative height of the two accent peaks, relative height seems a less reliable cue in the Background/Answer order (10a), and it has been proposed (e.g. Pierrehumbert and Hirschberg 1990: 296; Steedman 2000) that the distinction is also based on a phonological difference between two accent types – H* signalling 'Answer' (what Steedman calls *rheme*) and L+H* signalling 'Background' (what Steedman calls *theme*). However, experimental investigation by Calhoun (2006) reached a similar conclusion to Braun's study of the German hat pattern: there was no phonetic evidence for a categorical distinction between two accent types. In one part of her study, a trained speaker produced multiple instances of the distinction to the satisfaction of trained listeners; in another part, similar to Braun's study, naive speakers read sentences in contexts that forced the appropriate pragmatic interpretation. Like Braun, Calhoun found only that there were multiple acoustic dimensions on which small average differences between the supposedly contrasting accent types could sometimes be established. Calhoun did not directly study expert transcribers' treatment of her test utterances, but in this connection it is worth

noting that the accents in the *Anna/Manny* contour have been variably transcribed in the literature. Both accents were treated as H* by Pierrehumbert 1980; the specific suggestion that L+H* goes with the 'background' and H* with the 'answer' appears to originate with Pierrehumbert and Hirschberg 1990, and is fully developed by Steedman 2000.[6]

The two cases just discussed involve pragmatic distinctions that are notoriously difficult to think about: the literature on theme, topic, contrast, and givenness goes back many decades and there is still no clear consensus on how these concepts should be used. In particular, there is no agreement on whether the pragmatic concepts involved are categorical (e.g. theme vs rheme; Steedman 2000) or gradient (e.g. degrees of givenness; Prince 1981). However, there is at least one other instance of an intonational distinction in English that illustrates the same phonological problem – one category or two? – but that involves a more clearly categorical pragmatic distinction, namely the difference between statement and question. The phonological distinction at issue is the difference between what in a traditional British description was referred to as the distinction between 'low-rise' and 'high-rise' nuclear tones, and in AM analyses may be represented by the difference between L*_H_H% and H*_H_H%. This difference takes on special significance in the study of so-called HRT statements (High Rising Terminals; section 3.2.3.3), where it has been claimed (Fletcher and Harrington 2001; McGregor 2005) that the distinction between L* and H* can be used to signal the difference between statement and question and/or to convey other interactional differences. (A related claim is made by Warren 2005, but his phonological interpretation is somewhat different.) However, in both these studies the phonetic evidence for the supposed distinction between L* and H* is based on small average acoustic differences between HRT contours used in two pragmatically defined ways, not two distinct clusters of phonetic properties that define two phonological types. The similarity to Braun's and Calhoun's findings is clear.

6 In this connection it is worth mentioning a finding put forth by Dainora (2006) as evidence for a categorical distinction between H* and L+H*. Dainora shows, for two independently definable contexts in a corpus of transcribed spontaneous speech, that there is a significant difference between the two contexts in the proportion of H* and L+H* transcriptions used by expert transcribers. She then argues on this basis that the supposed categorical difference of accent type must be real. This does not follow. If there is a single category of rising accents with systematically conditioned variation along a continuum of realisations (which is suggested by both Braun's and Calhoun's work), and if the transcription system provides distinct labels for the two ends of the realisation continuum, it would actually be rather surprising if expert transcribers did *not* differ in the way they label accents from two systematically distinct contexts.

I believe that all the cases just discussed reveal something fundamental about the sound-meaning mapping in intonation. They all seem to call for a way of talking about the relation between intonational phonology and intonational pragmatics that allows us to recognise the existence of rather coarse phonological categories – like HRT or hat pattern – while at the same time deriving categorically distinct pragmatic interpretations from gradient variation in the way a given category is realised phonetically. This may seem uncontroversial, and in some cases it is. Specifically, when the gradient phonetic dimension involved is pitch range and the difference of pragmatic interpretation involves a difference of attitude (e.g. Hirschberg and Ward's 1992 finding that wider pitch range applied to the L*+H_L_H% contour tended to push the interpretation from 'uncertainty' to 'incredulity'), hardly anyone would disagree that we are dealing with meaningful gradient variation in the realisation of single categories. But what I am suggesting here goes beyond this. I believe that, even in cases where the two possible interpretations seem sharply distinct (e.g. question vs statement, theme vs rheme), and even when the differences in phonetic realisation involve dimensions other than overall pitch range (e.g. earlier vs later alignment), we are still dealing with a small number of categories that are subject to meaningful gradient variation. That is, even though it seems clear that question and statement or theme and rheme are categorically distinct *as semantic/pragmatic notions*, the apparently distinct intonational patterns that signal these categories may be only the opposite ends of a continuum of phonetic realisation.

In fact, there is a contradiction in much of the work just discussed that knows that many authors implicitly think in such terms already. For example, Fletcher and Harrington 2001 and McGregor 2005 speak of a difference between two distinct accent types L* and H*, yet nevertheless treat both as part a global phenomenon of 'HRT'. If the difference between L* and H* really is a categorical phonological distinction, then in principle the difference between H*_H_H% and L*_H_H% should be just as great as between, say, H*_H_H% and H*_L_H%. That is, there is no theoretical basis for taking H*_H_H% and L*_H_H% together as constituting a single area of study. The intuition that says – rightly, in my view – that HRT is a coherent phenomenon finds no expression in a strict theoretical interpretation of the proposed phonological difference. Similar comments apply to any discussion that distinguishes between an early aligned and a later aligned rising accent while continuing to refer to the 'hat pattern' or 'bridge accent'. What seems to be needed is a Gussenhoven-style phonology of the sort discussed in section 4.2.1, with a small number of categories and a set of dimensions of meaningful gradient modification, rather

than a system like that of Pierrehumbert and Hirschberg, which is based on a larger inventory of categorically distinct phonological elements. But the details remain elusive, and for now we are still plagued by 'paralinguistic stalemates'.

4.3 Intonation and lexical uses of pitch

One important topic remains to be discussed: so far we have said nothing about intonation in languages that have lexical tone. To the extent that the AM theory claims to provide a universal framework for discussing intonation, it is obviously important to establish that the intonational phenomena observed in tone languages can be accommodated in the theory presented so far. In particular, it is important to consider the relation between tone and intonation, or (to use the terminology of chapter 1) between lexical and postlexical pitch features. The remainder of this chapter is a brief discussion of three specific topics in this general area.[7]

4.3.1 The unity of pitch phonology

Perhaps the most important point to make at the outset of the discussion is that the AM theory, by analysing intonation contours at one level of description as strings of tones, provides the basis for describing pitch phonology in all languages in the same terms. As we saw above (section 1.3.2), there is good reason to believe that the relationship between lexical and intonational features of F_0 need not be conceived of in terms of an overlay model, with lexical features superimposed on intonational trends. Rather, in at least some cases – as in the case of the Swedish word-accent distinction as analysed by Bruce (1977) – the intonational aspects of F_0 are themselves localised events in a string of tones, exactly like the lexical aspects. In effect, from the point of view of the phonology (as opposed to the lexicon or the grammar), a pitch contour *in any language* is simply a string of tones. The tones may be organised into postlexical pitch accents, as in English, or they may be specified in lexical representations, as in Chinese. But the pitch phonology of all languages involves sequential tonal structure consisting of events and transitions; there is no fundamental difference in this respect between languages with and without lexical tone.

7 Practical constraints have forced me to leave this section substantially the same as in the first edition of the book, and I have been unable to take proper account of more recent work, especially an important paper by Hyman (2006). I would particularly urge the reader to regard the 2 × 2 set of distinctions in table 4.1 as a bare minimum: there is much more to discover about 'word-prosodic typology' than my preliminary conjectures here. Cf. also Remijsen 2002; Donohue 1997.

Given this view, the conspicuous phonetic difference between the pitch contours of Chinese and English, for example, is simply a *consequence* of the functional difference, and involves no essential difference of phonological type. More 'happens' in pitch contours in Chinese, because the lexical tones occur at nearly every syllable and the transitions between them span only milliseconds, whereas in English the pitch accents occur mostly only on prominent words, and the transitions may span several syllables. But the specifications are the same kind of phonological entity regardless of their function, and transitions are the same kind of phonetic phenomenon irrespective of their length. There is no need to assume that tone languages involve an essentially different layer of phonological structure.

On the other hand, it may be useful to coin the term *core tones* to distinguish them from edge tones in a cross-linguistically valid way. For example, Japanese clearly has both pitch accents and edge tones; the former are associated with lexically accented syllables, while the latter occur at the left edge – and in some cases also the right edge – of prosodic domains of various types. The edge tones act as signals of phrasing and in some cases also cue the difference between statements and questions, while the pitch accents are lexically specified and only ever involve the tonal pattern HL (see Pierrehumbert and Beckman 1988 for a full discussion). In many respects, then, the structure of the pitch contour is exactly like the structure of the contour in English: the two languages share a clear distinction between the F_0 events that make up the body of the contour and those at the contour's edges. In English, however, both types of features have a sentence-level or postlexical function; while in Japanese, the accentual elements are lexically specified, and only the edge tones are postlexical. By emphasising the structural difference between core tones and edge tones, rather than the functional difference between lexically specified tones and intonationally specified ones, we emphasise the common features of tonal phonology – the phonology of F_0 – in all languages.

From this perspective, the fundamental difference between languages with and without lexical tone is primarily a matter of how the tonal specifications come to be where they are. Core tones in Chinese are part of the phonological shape of morphemes, whereas in English they signal intonational 'sentence accent' on selected words of a phrase, yet edge tones in both languages are intonational. In Tokyo Japanese, the core tones signal lexically specified accent, while the edge tones are intonational, but in Osaka Japanese some of the edge tones are lexically specified as well (McCawley 1978; Kori 1987). In Yoruba, meanwhile, there may not even be edge tones (or, if there are, they

are completely predictable from the lexical specification of the last core tone).[8] Conversely, in some languages there appear to be no core tones at all, only edge tones; this is suggested by Rialland and Robert's description of Wolof (2001) and Lindström and Remijsen's description of Kuot (2005). All these different typological possibilities can readily be accommodated in the descriptive framework made possible by the AM phonological view.

4.3.2 *Tone–intonation interactions*

At least three kinds of phenomena have been cited as intonational features in tone languages. These are overall expansion or contraction of pitch range to express emotions, and/or to express intonational distinctions like question versus statement or incompleteness versus completeness; modification of specific tones, especially at the ends of phrases or utterances, to signal distinctions like question versus statement; and modification of overall contour shapes to signify certain intonational messages, again, in particular, question versus statement or incomplete versus complete. The first two of these have straightforward explanations in terms of the theory presented so far, and I will argue that it is probably also possible to explain the third type – overall contour modifications – in terms that are compatible with the idea of sequential tonal phonology.

First let us consider overall pitch range modification. In the view developed here, this is a paralinguistic effect that is orthogonal to the specification of the string of tones. To take an idealised example, a tone sequence M ... H ... H ... L ... M might be realised in different ways depending on the overall level of emotion of the speaker, as shown in figure 4.2.

This kind of pitch range effect occurs in exactly the same way in non-tonal languages: in English as in Yoruba we can equally well distinguish a 'normal' and 'expanded pitch range' version of something that is in other respects the same tune. That is, the relation between paralinguistic pitch range modification and linguistic tonal specification is identical in languages with and without lexical tone. Consequently, the existence of paralinguistic pitch range modification does not constitute evidence against the AM theory, or indeed against any phonological theory of intonational structure: whatever description we give of pitch range modification, it will apply in the same way in all languages.

The second 'intonational' phenomenon in tone languages – phrase-final tone modification – is similarly unproblematical for the AM theory. It has frequently

8 In Yoruba, final lexical L tones are realised with a fall to the bottom of the speaker's range (L edge tone?), while final H and M tones are realised as sustained levels or very slight rises (H edge tone? absence of edge tone?), but there is no tonal difference between statements and questions, or any other obvious intonational effect of sentence-level pragmatics.

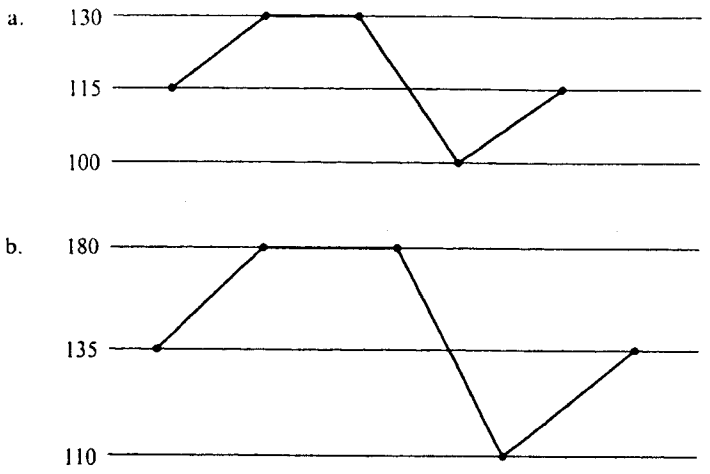


Figure 4.2 *Idealised pitch contours for a hypothetical tone sequence M ... H ... H ... L ... M, spoken in (a) normal and (b) expanded range.*

been reported that languages with lexical tone 'modify' the lexically specified tone at the end of a phrase or utterance to signal various utterance-level or postlexical effects (e.g. Chang 1958 for Chengdu Chinese; Abramson 1962 for Thai; and Lindsey 1985 for Hausa). For example, Chang (1958, reprinted in Bolinger 1972: 408f.) reports the following 'perturbations' of the tone of the final syllable in questions ('naming tone' refers to the citation form):

> *Toneme I* (naming tone: high-rising) remains high rising and often ends higher than usual.
> *Toneme II* (naming tone: low-falling) becomes low level.
> *Toneme III* (naming tone: high-falling) becomes high level.
> *Toneme IV* (naming tone: low-falling–rising) becomes low rising.

Chang herself refers to these changes as the result of superimposing the sentence melody on the sentence as a whole – that is, in terms of an overlay model – but given the fact that the perturbations are restricted to the final syllable, it seems equally plausible to regard them as being the result of associating the final syllable with a sequence of the lexical tone and a high boundary tone. That is, the basic shape of the contour is determined by the lexically specified tones on the morphemes that happen to make up an utterance; at the edge of the contour there can be additional tones that affect the realisation of the final lexically specified tone or tones.

This brings us to the third and (from the point of view of the AM theory) potentially most problematic intonational phenomenon in tone

languages, namely the reported occurrence of meaningfully distinct overall contour shapes on which the lexical tones are superimposed. This is obviously the kind of evidence that would support Bolinger's wave metaphor and the 'overlay' view discussed in section 1.3.2. If we wish to maintain the view developed so far – that contour shapes are specified only in terms of local phonological events – then we need to examine this evidence carefully to see whether it can be accommodated.

There are three fairly well-established types of cases that have been interpreted as evidence for the direct specification of overall shapes and therefore for the overlay view. These are:

(1) suspension of downdrift in questions;
(2) marking the location of focus;
(3) signalling at least a three-way intonational distinction (yes–no question versus WH-question versus statement).

Let us consider each of these in turn.

The suspension of downdrift in questions has been demonstrated for Hausa by Lindau (1986) and Inkelas and Leben (1990). In Hausa, as in many African languages, the second H in a sequence H ... L ... H is normally realised at a lower level than the first H. Moreover, this effect can be iterated, so that in a sequence H ... L ... H ... L ... H ... L ... H, each H tone is realised at a progressively lower level. This overall downward trend is clearly phonologically conditioned, however: in non-alternating sequences of tones like H ... H ... H ... H ... H, there is only a very slight 'declination', so that the slope of a 'topline' drawn through the H tones is much less steep than in the case of alternating sequences.

Now, this downdrift effect does not apply in questions. In a sequence H ... L ... H ... L ... H ... L ... H in a Hausa question, the H tones are all realised at approximately the same level, though there may be a slight amount of 'declination'. Obviously this means that a postlexical or intonational factor influences the realisation of the lexically specified tonal string. However, it does so in a way that is entirely in keeping with the idea of sequential phonological structure: rather than operating in terms of some overall slope appropriate for questions or for statements (which is what we might expect from, for example, Grønnum's description of Danish; see figure 1.5), it is clear that the Hausa intonation patterns are produced *locally*, by the operation or non-operation of the downstep rule at successive points in a sequence of tones (see section 2.4). The slope of a line through a statement contour is very much dependent on the lexical tonal makeup of the contour, and *cannot* be specified as a global shape

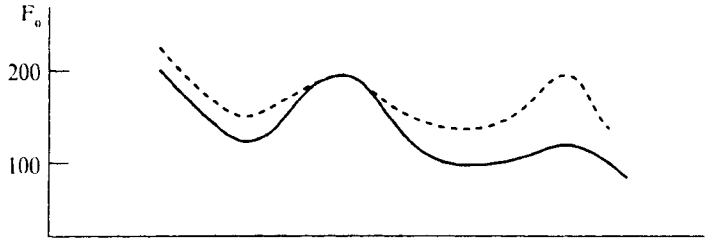


Figure 4.3 *Overall contour shapes signalling location of focus in Chinese. The utterance in both cases is* Sòng Yán mài niúròu *('Song Yan sells beef'), spoken by the same speaker. The dashed line shows the contour when the focus is late in the utterance and the pitch range remains more or less constant throughout; the solid line shows the contour when the focus is on the subject* Sòng Yán *and the pitch range narrows abruptly on the predicate* mài niúròu. *After Gårding 1987.*

applying to statements. There *is* an interaction between intonation and lexical tone, but it must be described in terms of local phonological events. This is exactly what the AM theory would lead us to expect.

The second case is more problematical, but still not beyond the reach of the AM view. This is the reported use of overall contour shapes to signal the location of focus. For example, Gårding (1987) suggested that 'early focus' sentences in Chinese have a 'grid' (cf. figure 1.4 and accompanying text) that narrows abruptly after the focus, whereas for normal 'late focus' sentences the grid remains wide, as shown in figure 4.3. The same claim has been put forward by Xu (2005) as part of his newer and more sophisticated version of the overlay view (discussed in section 1.3.2).

The AM interpretation of such cases will have to involve modifications of pitch range for emphasis. This can be related to pitch range modification for emotional signalling, discussed above. The difference is that in the case of emotional modification of pitch range, the modification affects an entire utterance, where in the case of focus and emphasis it appears that the modification of pitch range can take place during the course of the utterance. Thus the early focus sentences would have an idealised 'tonal space' (see section 2.3.4), as in figure 4.4. The normal late focus sentences, by contrast, would keep the tonal space unmodified throughout the utterance. As with the suspension of downdrift, we must obviously allow that in this case intonational or postlexical factors affect the realisation of the tones, but once again the intonational effects can be described as *occurring at a specific point or points in the utterance*

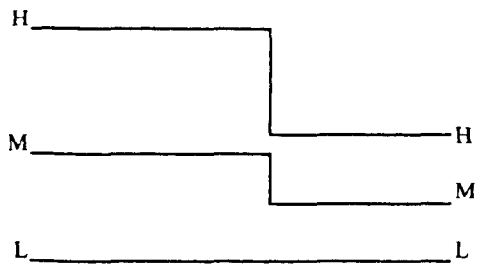


Figure 4.4 *Idealised 'tonal space' for modelling focus effects on pitch realisation in tone languages. The abrupt narrowing of the tonal space marks the end of the focus constituent. See text for more detail.*

rather than involving a contour shape that has to be specified as such in the phonology.

The third case is the most difficult, namely the possibility that overall contours with distinct shapes convey distinct intonational meanings. The most serious evidence for this comes from a proposal by Shen (1990) for Chinese, which has been applied to Kipare by Herman (1996). Shen claims that there are at least three distinct intonational tunes, which she calls Tunes I, II, and III. These are used, according to Shen, for declaratives (Tune I), yes–no questions (Tune II), and WH-questions and alternative ('is it A or B') questions (Tune III). Herman suggests that similar tune types are used in Kipare, for declaratives, incredulous questions, and ordinary yes–no questions respectively. Shen's graphic portrayal of the three tunes is shown in figure 4.5. [9]

In order to accommodate these in an AM theory, one of two things will be necessary. The simplest possibility would be to interpret the overall shapes as involving local pitch range specifications – that is, to show that they result from successive expansions or contractions of pitch range that occur at well-defined locations. This kind of interpretation is not inconsistent with what Shen says about the three tunes. Her description is as follows (1990: 26):

> *Tune I*: starting with a mid key, moving upward to a mid-high key at the highest peak, falling to a low register at the ending point.
> *Tune II*: starting with a mid-high key, moving upward to a high key at the highest peak, dropping, but not too low, ending in the high or mid-high register.
> *Tune III*: starting with a mid-high key, moving upward to a high key at the highest peak, stepping down and ending with a low key.

9 The numbering of the tunes in Shen's figure disagrees with that in her prose definitions. My usage here is based on her prose definitions.

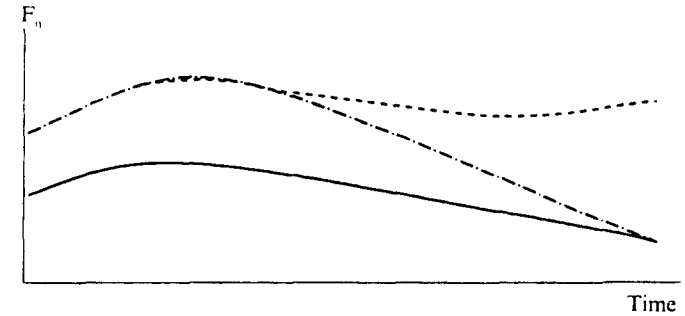


Figure 4.5 *Chinese sentence-level 'tunes', from Shen 1990. The three tunes are intended to indicate the approximate overall shape of the utterance contour in declaratives (Tune I, solid line), yes-no questions (Tune II, dashed line), and alternative and WH-questions (Tune III, dash-dot line).*

If we treat Shen's impressionistic descriptions of the different keys or registers as characterisations of the phonetic realisation of the tonal space, it is at least possible that we could show that the *changes* from one key to another happen at well-defined points in the contour. Herman's description of the three tunes is clearly consistent with this sort of interpretation: in her analysis, the Kipare equivalent of Tune I involves final lowering and no pitch range expansion; Tune II exhibits expanded pitch range without final lowering; and Tune III has a combination of expanded pitch range and final lowering. If such a description of these global tunes in Chinese were possible, it would be comparable to the treatment of the pitch range effects involved in early focus sentences, discussed just above.

If this interpretation turns out not to be possible and it is necessary to describe the three overall shapes as global gestures, then this finding could be made compatible with the AM theory by showing that analogous effects on pitch range are found in non-tonal languages as well. For example, we might seek to show that English distinguishes contours in which the pitch accent peaks are of moderate height and gradually decline (Tune I) from contours in which the peaks are of much greater height and decline very little (Tune II). This is essentially what Grønnum shows for Danish (cf. figure 1.5). Such an interpretation might seem like an important theoretical concession, and indeed it is; however, it is important to keep in mind that the question under consideration is *whether the phonological structure of pitch is comparable in all languages*. If it turned out that declination were meaningfully variable as such in all languages, then that fact would have to be built into our phonetic model.

But in all languages it would still be possible to distinguish sequential tonal specifications – English postlexical pitch accents or Chinese lexical tones – from pitch range modifications: that is, all languages would still require *both* tonal specifications and descriptions of how pitch range can be modified. Lexical tone would still not constitute an additional layer or special type of phonological specification, and the 'unity of pitch phonology' would remain unaffected.

4.3.3 *'Pitch accent' languages*

Finally we come to a third topic in the general area of tone–intonation interactions, namely the phonetic typology that Beckman (1986) has discussed under the title 'stress and non-stress accent'. Beckman's title unfortunately introduces a terminological confusion to which attention must be drawn from the outset. Her 'accent' might be defined as 'the special phonological treatment of a particular syllable or syllables in a word'. Given this definition, 'stress accent' and 'non-stress accent' are two ways in which the special treatment of the syllable can be manifested phonetically. 'Accent' is thus an abstract phonological feature specified in the lexicon; 'stress' is a phonetic description of one possible concrete realisation of the phonological abstraction. In what follows I will refer to Beckman's sense of accent as *lexical accent*.

As we have already noted (section 2.2.1), Beckman shows that in languages like English, lexical accent is regularly manifested by what we called stress, a cluster of phonetic properties that includes increased intensity and duration as well as various spectral correlates. *In addition*, stressed syllables are often accompanied by the major pitch movements we called pitch accent. Beckman compared English with Japanese, showing that in Japanese, lexical accent is marked only by pitch movement, and not at all by stress. On this basis she supports the traditional typological distinction between 'dynamic' and 'melodic' accent: dynamic or stress accent is lexical accent realised by stress (and often accompanied by pitch movement), while melodic or non-stress accent is realised always and only by pitch movement.

Because Beckman's distinction is based on a comparison between English and Japanese, however, there is a potential for confusion that I wish to mention here. In English, pitch features are only postlexical or intonational, while in Japanese pitch features are specified in the lexicon. However, that difference has nothing at all to do with the distinction between stress and non-stress accent. Stress versus non-stress accent is a *phonetic* typological dimension (is lexical accent manifested by stress or not?). Lexical versus postlexical specification of pitch features is a *phonological* or even morphological typological dimension (can pitch features be specified in the lexicon or not?). It is important not to

Table 4.1 *Proposed independence of lexical typology and phonetic typology in cross-linguistic uses of word-level prosodic features.*

		Phonetic typology	
		Stress accent	Non-stress accent
Lexical typology	Lexical pitch	example: Swedish	example: Japanese
	Postlexical pitch only	example: English	example: Bengali

assume that 'non-stress accent' is a matter of lexically specified pitch features, just because Japanese happens to work that way.

If the two typological dimensions are really independent, it ought to be possible to find four types of languages. In particular, we ought to be able to find languages that have lexically specified pitch features with stress accent, and languages that have only intonational pitch features and non-stress accent. I believe that this can be done: Swedish (like most of the European 'pitch accent languages') is an example of a language combining lexically specified pitch features with stress accent, while Bengali (and probably most of the languages of India) is an example of a language with non-stress accent and no lexical specification of pitch. This typological conjecture is summarised in table 4.1.

In this connection we should explicitly mention the appropriate analysis of the European 'pitch accent' languages like Swedish and Serbo-Croatian. To my knowledge, all such languages clearly have 'stress' in Beckman's phonetic sense, and at least some of them have the possibility (like English or Bengali, and unlike Japanese) of selecting different pitch accent types to convey different pragmatic meanings. (In the clearest such example, Serbo-Croatian has a different pitch accent type for use in questions – the 'reverse pattern' of Lehiste and Ivić (1980, 1986) – from the pattern used in statements and citation forms.) Where such languages differ from languages like English is in the fact that the choice of pitch accent type seems to be influenced by lexical considerations as well as postlexical ones. Thus Swedish distinguishes between Accent 1 words and Accent 2 words; Serbo-Croatian distinguishes falling and rising accents; and so on.

The correct tonal analysis of these languages is not altogether clear. There have been a number of suggestions for these languages that the accentual contour itself is constant (say, HL), and that what varies is the alignment of the accentual contour with the segments on the accented word. Perhaps the most noteworthy of these analyses is Bruce's work on Swedish, which we have discussed extensively (sections 1.3.3 and 2.3.4), but similar analyses have been

given for other languages, such as Browne and McCawley's (1965) analysis of Serbo-Croatian accent. However, a variety of evidence now seems to point to the conclusion that the pitch accent configurations can involve a combination of lexical and postlexical specifications. For example, Inkelas and Zec (1988) analyse the Serbo-Croatian accent distinction as based on the location of a prelinked lexically specified H tone: the falling accents have a lexically associated H tone on a word-initial syllable, and the rising accents have a lexically associated H on a non-initial syllable. The latter are then subject to leftward spreading, and in some cases to a postlexical final-lowering rule, while the former are not. In a series of descriptions of the word accent distinctions of the Central Franconian (Rhineland) dialects of Dutch and German, Gussenhoven and his colleagues (e.g. Gussenhoven and van der Vliet 1999; Gussenhoven 2000) take a similar approach: for most of these dialects they argue that the basic lexical accent distinction is between the presence and absence of a lexical H tone on the stressed syllable, and the actual spoken contours depend on the interaction of any lexical tones with other tones that are assigned postlexically.

Such analyses are broadly consistent with evidence about what happens to the Swedish word accents when they are emphasised (Fant and Kruckenberg 1994). Recall that Swedish citation-form contours comprise a sequence of an accentual fall and a 'phrase accent' peak. For Bruce (1977), the fall is HL and the phrase accent is H, and the only difference between Accent 1 and Accent 2 is that the HL is aligned earlier with Accent 1. However, by asking speakers to pronounce the word accents with different degrees of overall emphasis, Fant and Kruckenberg showed that the L in Accent 2 is not the same as the L in Accent 1. When the overall emphasis was increased on Accent 1, all three turning points – accentual H and L and phrasal H – increased in F_0. By contrast, when overall emphasis was increased on Accent 2, the H tones were raised and the L tone was lowered. This is shown in figure 4.6. While it is by no means clear exactly how to translate these findings into lexical representations, it is not implausible to suggest that the L tone is somehow more 'real' in Accent 2 than in Accent 1. If we interpret 'real' as meaning lexically specified, then the Fant and Kruckenberg data could be seen as consistent with the analyses of the Franconian accent systems devised by Gussenhoven and his colleagues.

Once again, we see here the importance of assuming what I called the unity of pitch phonology. The AM approaches to pitch accent sketched here draw no fundamental distinction between tones that are assigned in the lexicon and tones that are assigned or modified postlexically. Earlier, I suggested a distinction for

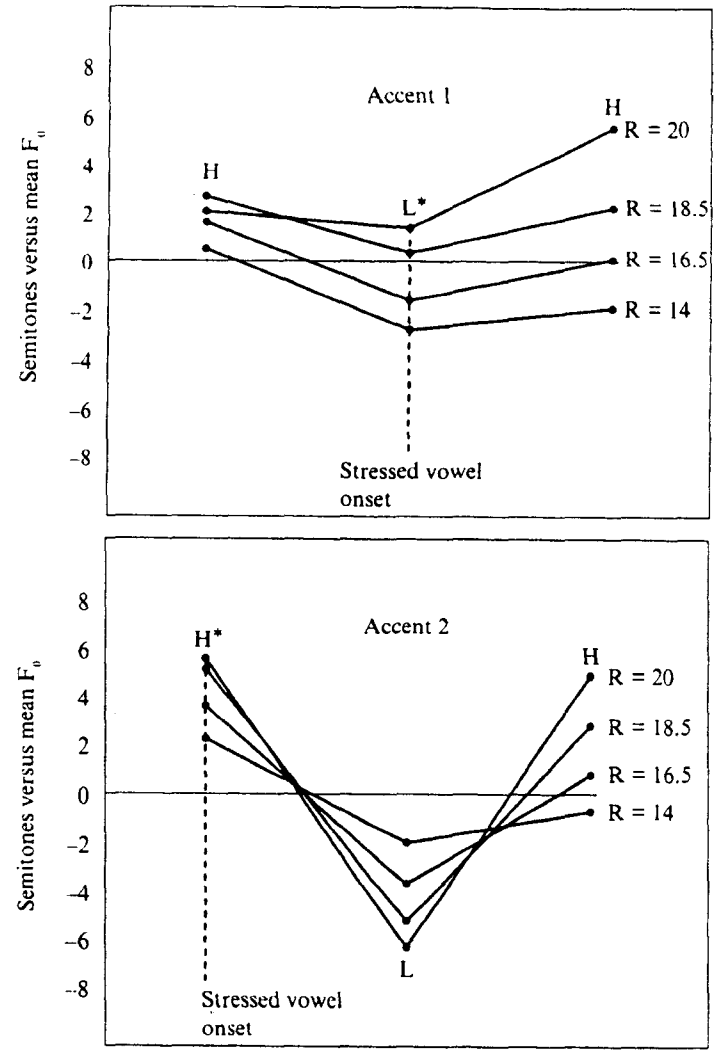


Figure 4.6 *Effect of increasing emphasis on the pitch configurations of the Swedish word accents. In Accent 1 the accentual H and L and the phrase tone H are all raised with increasing emphasis. In Accent 2 the L is* lowered *and only the H tones are raised. From Fant and Kruckenberg 1994.*

Chinese and Japanese between lexical core tones and postlexical edge tones, but in the case of the European pitch accent languages it appears that lexical and postlexical tones may be interspersed. Phonetically the distinction between the two is irrelevant: for all pitch contours in all languages, the input to phonetic realisation is a string of tones.

5 *Intonational phonetics: alignment and scaling of tonal targets*

Given the autosegmental understanding that the pitch contour of an utterance is the realisation of a string of tones, an obvious question for phonetic research is how these tones are realised. The phonetic manifestation of the tones is generally conceived of as some sort of tonal target; the principal phonetic dimensions on which tonal targets can vary are their pitch level – often referred to as *scaling* – and their temporal coordination with the consonants and vowels of the segmental string – often referred to as *alignment*. This autosegmental perspective has been extremely productive of instrumental phonetic research since the 1980s; and on the whole this research has shown that both scaling and alignment are highly lawful and can be systematically influenced by a range of phonetic and phonological effects. This work, together with some of its phonological implications, is reviewed in this chapter.

5.1 Alignment of tonal targets with the segmental string

5.1.1 Alignment as a phonetic dimension

At least since the early work of the IPO researchers in the 1970s, it has been clear that intonational distinctions can be conveyed by differences in the way pitch movements are aligned with the segmental string. For example, the IPO analysis of Dutch intonation distinguished between Type 1 and Type 3 rises, both of which were said to be 'prominence-lending' pitch movements (i.e. pitch accents); the difference between them is that the Type 1 rise is aligned early in the prominent syllable, while the Type 3 rise is aligned somewhat later. This distinction was matched by a distinction between Type A and Type C falls – Type A being relatively early and Type C being relatively late. These rises and falls can be combined to create 1A and 3C rising–falling accents, the 1A being a normal declarative 'pointed hat' accent, and the 3C conveying an additional nuance. The idea that essentially similar pitch movements can be aligned with the segmental string in meaningfully different ways is thus clearly stated in the

IPO work on Dutch; it finds its way into Odé's IPO-style description of Russian (1989) as well.

The same general idea was apparently independently developed by researchers on the Scandinavian word-accent distinction. Haugen and Joos 1952 came very close to stating that the essential difference between Accent 1 and Accent 2 in Norwegian is a matter of different alignment of the accentual pitch movement with the accented vowel. The alignment idea was made explicit and extensively developed in Gösta Bruce's thesis (1977), in ways that made important contributions to the establishment of the AM approach to pitch phonology. Recall (section 1.3.3) that the essence of the Swedish word-accent distinction, according to Bruce, is the way the H of the accent contour – the F_0 maximum – is aligned relative to the stressed syllable. Specifically, Bruce found that the distinction between Accent 1 and Accent 2 is reliably cued by the timing of a high peak early in the accentual configuration – before the beginning of the accented syllable for Accent 1 and just after the onset of the accented vowel for Accent 2. Normally this high peak is followed by a fall, but Bruce found that the fall can be abbreviated or lost altogether under certain conditions; only the high and its alignment are invariant.

Earlier accounts of the Swedish word-accent distinction – with the partial exception of Haugen and Joos 1952 – had concentrated on the pitch contour *of the accented syllable itself*, and researchers were puzzled by the extreme variability of the syllable contours they observed (see e.g. Hadding-Koch 1961). Bruce's insight was to see that the F_0 contour on the accented syllable is merely a consequence of which tones occur and how they are aligned with the segmental string. If H is aligned at the beginning of a syllable, the syllable will have a falling contour, whereas if the H is aligned at the end of a syllable, the syllable will show a rising contour. On the basis of this insight, Bruce and Gårding (1978) went on to produce a prosodic typology of Swedish word accents, showing that the accentual distinction in all varieties that have it is signalled by the earlier alignment of the peak of Accent 1, even though the peaks of both accents may be aligned relatively earlier or later in different dialects, and even though there may be other concomitant phonetic or phonological differences.

Since Bruce's thesis, several experiments have confirmed that differences of alignment can signal quasi-categorical pitch distinctions in several languages. The first such study seems to have been carried out by Kohler (1987). Kohler manipulated the alignment of a rising–falling pitch movement relative to the stressed syllable *-lo-* of the German sentence *Sie hat ja gelogen* ('She's been lying', lit. she has yes lied) and presented the stimuli to a large number of listeners in a classical categorical perception study (cf. e.g. Repp 1984). The

results provided evidence for a sharp discontinuity in the stimulus continuum between an earlier alignment category (with peak F_0 during the unstressed syllable *ge-*, i.e. in AM terms probably an H+!H* accent) and a later alignment category (with peak F_0 during the nuclear accented syllable *-lo-*, i.e. in AM terms an H* or L+H* accent). Kohler suggested that the perceptual boundary corresponds to a semantic/pragmatic difference between early aligned given (which he calls 'established') and later aligned new; this is confirmed by Baumann's observations (2006) on the use of downstepped accents in German (cf. chapter 2, footnote 11 on p. 78). Kohler also suggested that there is actually a three-way distinction given-new-emphatic, with 'emphatic' being aligned later than ordinary 'new' (cf. the discussion in section 4.2.2), but this was not part of his experimental study. The idea of a three-way or even four-way distinction has more recently been taken up by Kohler 2005 and Niebuhr 2007.

Several other studies of alignment, on a range of European languages and with a range of methodologies, have obtained results similar to Kohler's. For example, Pierrehumbert and Steele (1989) manipulated the rising–falling–rising pitch contour of the naturally spoken American English phrase *Only a millionaire* to yield a continuum of contours in which the first (accentual) rise was aligned earlier or later relative to the lexically stressed syllable *mil-*, and presented the manipulated utterances to listeners who were asked to mimic the intonation. They predicted that the listeners' productions would not faithfully reproduce the continuum, but would fall into two alignment categories reflecting the difference between an L+H* (early) and an L*+H (late) rise. For at least some of the participants in the experiment this prediction was confirmed. The same general technique has been used in a larger study by Dilley and Brown (2007), with similar results, though Dilley and Brown's methods – and their theoretical interpretation – focus on the relative F_0 level of the accented syllable and the following syllable, rather than on alignment of a specific peak.

In a more traditional perception study, Rietveld and Gussenhoven (1995) created stimulus continua in which the nuclear falling pitch movement of a Dutch 'flat hat' contour (cf. section 1.2.1) was aligned earlier or later relative to the nuclear accented syllable (which was one of a set of monosyllabic nonsense words). They then asked listeners to report whether the accented syllable sounded high (later alignment of the fall) or low (earlier alignment). Listeners' responses showed typical s-shaped curves, reflecting categorical shifts between 'low' and 'high' at clearly localisable places in the alignment continuum. About the same time, Verhoeven (1994) reported comparably categorical findings for the Dutch hat pattern's rising prenuclear accent. Experiments by D'Imperio and House (1997), using alignment continua

of rising–falling nuclear accent contours in Neapolitan Italian, showed that listeners responded categorically in their judgements of whether test sentences were intended as statements (early alignment) or questions (late alignment); and a very similar phonetic difference between statements and questions in Russian has been reported by Makarova (2007).

From the work just reviewed we can conclude that phonetic continua of F_0 alignment can trigger essentially categorical responses from listeners, and that it is appropriate to regard the alignment of pitch movements relative to the segmental string as a phonetic dimension on which phonological distinctions can be based. Note that this conclusion is essentially theory-neutral, in the sense that it does not in any way presuppose AM intonational phonology: it is *consistent* with the AM idea that pitch contours are composed of strings of tonal targets aligned in a clearly defined way with the segmental string, but it is compatible with other views as well. Many of the findings just summarised, like the methodology of some of the experiments on which they are based, are expressed in terms of the alignment of rises and falls, not F_0 turning points or tonal targets, and at the very least they admit of different theoretical interpretations. But regardless of whether alignment data ultimately support the AM approach to intonational phonology, there is a real phonetic phenomenon at work here, which any understanding of intonation is going to have to take into account.

5.1.2 *Segmental anchoring*

While the facts discussed in the preceding subsection may be relatively theory-neutral, better evidence for a specifically AM interpretation of F_0 segment alignment comes from a large number of controlled 'production' studies of the acoustic details of intonation contours. Like Bruce's early work, many of these studies have been based on the alignment of F_0 maxima and minima – 'turning points' – which, as we saw in section 4.1.2, are readily interpretable as the phonetic reflexes of phonological tones. These studies have repeatedly found that the alignment of F_0 turning points is extremely consistent and predictable.

Such consistencies were apparently first noted for English by Ashby (1978). Ashby was investigating the acoustic correlates of the 'high-fall' and 'low-rise' nuclear accent configurations of the traditional British analysis of English intonation. He had three phonetically trained speakers read a long randomised list of sentences, with no instructions on how to produce them except to specify one or the other nuclear accent type. Ashby found to his surprise that, while there was a good deal of variation in the prenuclear stretches of the sentences, the nuclear accent realisations were extremely consistent. For the high-fall contours, for each speaker considered separately, both the scaling and the alignment of the peak showed very little variation: for example, one male speaker had peaks of about 200 Hz aligned about 35 ms after the onset of the vowel. The final low of the high-fall contours on sentence-final nuclear monosyllables was also extremely constant: the same male speaker reached a low of about 100 Hz 150 ms after the vowel onset. The low-rise contours were characterised by a low, very slightly falling 'plateau whose length is a linear function of the total voicing time after vowel onset' (p. 334), followed abruptly by a rise of fixed slope (6.7 octaves/sec.) that continued to the end of voicing.

A decade after Ashby, the first AM-inspired instrumental work demonstrated that the alignment of turning points can be predictably variable as well as predictably stable. Two studies around this time (Silverman and Pierrehumbert 1990; Prieto, van Santen, and Hirschberg 1995) investigated the systematic influences of prosodic context – specifically, the proximity to other pitch accents ('stress clash') and prosodic boundaries – on the alignment of accentual pitch peaks. Silverman and Pierrehumbert manipulated the distance between a prenuclear high accent and the beginning of the following (nuclear-accented) word, using made-up names like *Mom Lemm*, *Mama Lemm*, *Mamalie LeMann*. They showed that the peak of the prenuclear accent is aligned proportionally earlier in the syllable the closer it is to the following word – that is, the upcoming word boundary and/or upcoming nuclear accent seem to exert some backward time pressure on the alignment of prenuclear peak. (They also found effects of speech rate, with the prenuclear peaks being aligned later in the syllable when the speech rate was increased.) In much the same vein, but with a rather more complex set of variables, Prieto, van Santen and Hirschberg (1995) studied the alignment of accentual high peaks in Mexican Spanish. Like Silverman and Pierrehumbert, they found extremely consistent interacting effects on alignment, based on the position of the accented syllable in the test word (viz. lexical stress on initial, medial, or final syllable); the position of the test word relative to upcoming prosodic boundaries (final in sentence, final in a phrase preceding a separate phrase, or followed by another word in the same phrase); and the position of the stressed syllable of the following word if there was one. For example, they found that accent peaks in word-final stressed syllables were aligned earlier, relative to the syllable, than accent peaks in initial or medial stressed syllables, and that accent peaks in phrase-final words (= nuclear accents) were aligned earlier than accent peaks in phrase-medial words (= prenuclear accents). Again, these results strongly suggest some sort of backward time pressure from the upcoming prosodic boundary.

These findings formed the background for a study of pitch peak alignment in Greek prenuclear accents by Arvaniti, Ladd, and Mennen (1998), which

discovered the phenomenon that has come to be known as *segmental anchoring*. Arvaniti and Ladd (1995) had already shown that the beginning of the rise in prenuclear accents in Greek coincides quite precisely with the beginning of the accented syllable (specifically, it seems to be aligned on average a few milliseconds before the beginning of the onset consonant). They also knew that the end of the rise – the peak of the prenuclear accent – shows rather variable alignment and seems to be affected by following boundaries and accents in much the same way as had been shown by Silverman and Pierrehumbert, and by Prieto, van Santen, and Hirschberg, for English and Spanish. To shed more light on the nature of these effects, Arvaniti, Ladd, and Mennen distinguished between what we might call a 'phonological' definition of time pressure (e.g. number of intervening syllables between the affected pitch accent and the following boundary or accent) and a 'phonetic' definition (e.g. actual duration of the time interval between the two), and concentrated their attention on the latter by keeping the syllable count constant. In all their materials the test syllable was always antepenultimate in the word, and there were always three unstressed syllables between the test syllable and the following accented syllable. They manipulated duration by using inherently shorter (e.g. /-tikoto-/) and longer (e.g. /-ɣlosiksa-/) sequences of intervening unstressed syllables.

Under these conditions, unlike Silverman and Pierrehumbert or Prieto, van Santen, and Hirschberg, they found no systematic effect of time pressure on the alignment of the peak. Instead, to their surprise, they observed that the peak was always aligned near the beginning of the unstressed syllable immediately following the test syllable, while at the same time replicating the finding of Arvaniti and Ladd (1995) that the preceding valley – the beginning of the accentual rise – is always aligned at or just before the onset of the accented syllable. In short, their results showed the kind of consistency in alignment that Ashby had reported in 1978.[1] Further experiments confirmed the initial observation. Importantly, the further experiments showed that the beginning and end of the accentual rise are aligned in the same way regardless of the duration of the segmentally defined interval between the specified alignment points. The temporal interval between the beginning of the accented syllable and the following unstressed vowel is greatly variable, depending on the segmental sequences

1 It should be emphasised that this statement applies specifically to the experimental situation where there are three intervening syllables between the test syllable and the following accent. Arvaniti, Ladd, and Mennen also found time pressure effects of the sort reported by Silverman and Pierrehumbert, and Prieto, van Santen, and Hirschberg, but only by reducing the number of intervening syllables; for some speakers, effects on alignment arose only when the number of intervening syllables was reduced to zero.

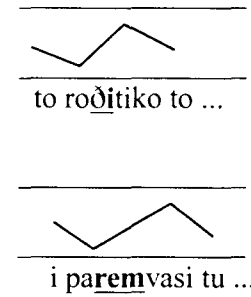


Figure 5.1. *Idealised pitch contours illustrating the 'segmental anchoring' found by Arvaniti, Ladd, and Mennen 1998. The F_0 turning points at the beginning and the end of the accentual rise are aligned in time with, respectively, the beginning of the accented syllable and the following unstressed vowel. This means that the duration of the rise depends on the duration of the interval between those two segmental landmarks. Since the F_0 turning points are scaled at roughly constant levels for any given speaker, the slope of the rise is also variable.*

involved, but the alignment remains unaffected. This obviously entails that the rise does not have a fixed duration. Moreover, given that the scaling (F_0 level) of both the beginning and the end of the rise was relatively constant for each speaker, the slope of the rise is not fixed either, but is adjusted according to the amount of time available for the rise (see figure 5.1). Instead, it appears that what is phonetically invariant about Greek prenuclear accents is the alignment of the beginning and ending points of the rise, not the slope or duration of the rise as a whole.

Arvaniti, Ladd, and Mennen argued that this finding provides strong support for the AM assumption that pitch movements are not the fundamental building blocks of intonation, but are merely transitions from one tonal target to another. Most previous acoustic models of F_0 had been based on pitch movements with standard slopes (e.g. 't Hart 1979) or standard durations (e.g. Fujisaki 1983), in keeping with the assumption that such movements are the elements of intonation. In fact, at least one experimental study (Caspers and van Heuven 1993) was explicitly aimed at determining whether slope or duration would take precedence when pitch movements were realised under time pressure, and had yielded results that could best be described as inconclusive, though it also observed the consistent alignment of certain turning points. Arvaniti, Ladd, and Mennen suggested that these earlier phonetic models implicitly adopted the 'configurations' point of view in the levels-vs-configurations debate

(section 2.3.1), and that a more adequate phonetic description would result from measuring the properties of F_0 targets rather than F_0 movements.

Arvaniti, Ladd, and Mennen's finding of segmental anchoring has since been replicated many times in many languages. For example, Ladd *et al.* (1999) – who coined the term *segmental anchoring* – showed that English rising prenuclear accents remain anchored to segmental landmarks regardless of speech rate: slope and duration are adjusted to keep the beginning and the end of the accentual F_0 rise aligned with their respective segmental anchors, as segment durations decrease or increase with rate. Further findings of consistent segmental anchoring include Dilley, Ladd, and Schepman (2005) on English, Ishihara (2003) on Japanese, and Prieto (2005) on Catalan, all of which show clearly that slope and duration of a pitch movement are not constant, but are adjusted according to the distance between the segmental anchors of the beginning and the end of the movement. A number of Xu's papers on the coordination of tone and segments in Chinese report findings reminiscent of segmental anchoring, in particular the fact that the pitch peak at the end of a rising tone preceding a low or falling tone is precisely aligned with the end of the syllable that is the phonological host of the rising tone (Xu 1998).

Arvaniti, Ladd, and Mennen (1998) initially suggested that segmental anchoring could be seen as a direct reflection of the autosegmental association of specific phonological tones to specific points in segmental structure. For example, we might explain the alignment of the Greek rising prenuclear accent by saying that in the intonational phonology there is an L tone associated with the initial syllable boundary, and an H tone associated with the following unstressed vowel, as shown in figure 5.2. This kind of interpretation was given an important boost by Ladd, Mennen, and Schepman (2000), who showed that the location of the segmental anchor points is sensitive to phonological characteristics of the segmental string. Specifically, they found that the beginning of prenuclear rising accents in Dutch is aligned at the beginning of the accented syllable, as in Greek, but the alignment of the end (i.e. peak) of the rise depends on whether the vowel is phonologically long or short (tense or lax): the peak accompanying a long vowel is aligned late in the vowel, but that accompanying a short vowel is late in the following consonant. They were able to show that this is not a mere phonetic effect of segment duration: Dutch 'long' /i/ and 'short' /ɪ/ are essentially identical in duration and differ only in vowel quality, yet a difference in alignment is still found.

On the basis of this finding, Ladd, Mennen, and Schepman suggested that the effect of phonological vowel length on alignment in Dutch could be explained

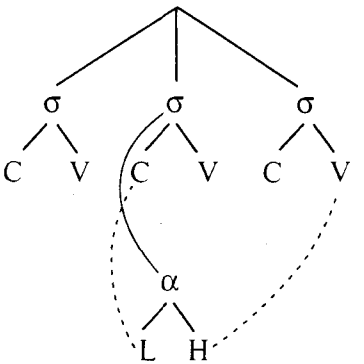


Figure 5.2. *A possible phonological interpretation of segmental anchoring as secondary association. The accentual configuration LH is primarily associated as a unit to the accented syllable as a unit, and the component tones L and H are then secondarily associated to, respectively, the left edge of the accented syllable and the following unaccented vowel.*

not in terms of phonological vowel length, but in terms of syllable structure. Their reasoning was as follows: in Dutch, a single intervocalic consonant following a long vowel is unambiguously the onset of the following syllable, while following a short vowel it may be interpreted as ambisyllabic (Schiller, Meyer, and Levelt 1997). The pattern of results observed in their study could therefore be interpreted as alignment of prenuclear peaks *at the end of the accented syllable*. In effect, Ladd, Mennen, and Schepman suggested that the effect of vowel length was indirect, and that their data could be explained phonologically if the accent were analysed as an L tone associated with the left syllable boundary and an H tone associated with the right.

Subsequent work by Schepman, Lickley, and Ladd (2006), however, suggests that the alignment effect reported by Ladd, Mennen, and Schepman 2000 may be directly conditioned by vowel length, and shows that it certainly cannot always be explained in terms of syllable structure. Schepman, Lickley, and Ladd found that the effect of vowel length on alignment is present with Dutch nuclear accents as it is with prenuclear accents, but in all cases the nuclear peak is aligned well within the accented vowel, and the expected effect of syllable structure in disyllabic nuclear-accented words is absent. Unfortunately, though, Ladd, Mennen, and Schepman's original syllable-based interpretation of segmental anchoring has already put down deep roots. A number of authors have taken the syllable-edge idea as somehow being a defining characteristic

of the 'segmental anchoring hypothesis'. For example, Prieto and Torreira (2007) find that prenuclear accent peaks in Spanish are aligned differently depending on whether the accented syllable has a coda consonant or not: because the alignment regularities they observe cannot be stated in terms of syllable edges, they suggest that this finding poses a challenge to the idea of segmental anchoring. Similar comments can be found in Xu's work (e.g. Xu and Sun 2002: 1441; Xu and Xu 2005: 163).

It therefore seems important to emphasise that segmental anchoring is not in the first instance a phonological hypothesis, but an empirical finding of a phonetic regularity. As a matter of observable, readily replicable fact, pitch movements are coordinated with the segmental string in such a way that the timing of their beginning and ending points can often be specified relative to landmarks in the segmental string itself – landmarks such as 'beginning of accented syllable', 'halfway through the consonant following the accented vowel', 'approximately 15 ms into the following unstressed vowel', and so on. Some of these segmental landmarks may be edges of phonological domains (e.g. beginning of accented syllable), but others (e.g. 15 ms into the unstressed vowel) are not. It is also observable and replicable that phonological conditioning factors may play a role in determining the details of the phonetic regularity. Ladd, Mennen, and Schepman (2000), as we saw, showed that alignment is different with long and short vowels in Dutch, and that the effect cannot be explained in terms of actual vowel duration;[2] Prieto and Torreira (2007) show that the details of segmental anchoring in Spanish are conditioned by syllable structure. But acknowledging the existence of phonological conditioning is not the same thing as treating segmental anchoring as a direct reflection of the phonological association of individual tones.

Further doubt is cast on a direct phonological representation of segmental anchoring by several studies that have found evidence of systematic differences in alignment between different languages or between different varieties of the same language. Just as Bruce and Gårding found that overall alignment can vary from one Swedish dialect area to another without affecting the generalisation that Accent 2 is aligned later than Accent 1, so other studies have found that the realisation of 'the same' contour in different dialects or different languages can vary along the dimension of alignment. For example, Atterer and Ladd (2004) reported that prenuclear accentual rises are later in Southern German than in Northern German, and later in German than in English; this finding has

2 A nearly identical effect of phonological vowel length on the alignment of prenuclear accent peaks has been demonstrated in Scottish Standard English by Ladd *et al.* (forthcoming).

since been replicated and extended by Mücke *et al.* (forthcoming). Arvaniti and Garding (2007) found a similar difference between Southern California (earlier aligning) and Minnesota (later aligning) speakers of American English. Similar findings are reported for different dialects of Irish by Dalton and Ní Chasaide (2005), and for different varieties of Basque Spanish by Elordieta and Calleja (2005). Ladd *et al.* (forthcoming) report that accentual peaks are aligned later in Scottish Standard English than in Southern British ('RP') English; they also report that English nuclear accent peaks are aligned earlier than in Dutch.[3] All these findings suggest that we are dealing with fine details of phonetic realisation, not distinct phonological representations: the differences reported are analogous to the range of phonetic realisations of the English short front vowels in *bid, bed, bad*, which vary considerably from dialect to dialect without affecting the set of phonological distinctions.

The foregoing discussion highlights the need for a terminological and conceptual distinction that I explicitly drew in Ladd 1983a, between *alignment* and *association*. Association, as in autosegmental phonology generally, is the abstract phonological property of 'belonging together' in some way. Alignment is a *phonetic* property, namely the relative timing of events in the F_0 contour and events in the segmental string. For example, more or less by definition, the word-accent contour on a given Swedish word is associated with the lexically stressed syllable, regardless of whether it is Accent 1 or Accent 2, while the phonetic difference between the two is a matter of the alignment of the word accent fall relative to the lexically stressed syllable. This means that the simple fact of (phonological) association entails no detailed predictions about (phonetic) alignment. Obviously, if an H tone is associated with a given prominent syllable, we may expect to find a peak of F_0 somewhere in the general vicinity of that syllable, not several syllables earlier or later, but the peak may be early in the syllable or late, and indeed it may be outside the temporal limits of the syllable altogether. Association should be kept as a phonological concept, and detailed differences of alignment should generally not be given a direct phonological representation.

3 Though these studies are based firmly on AM assumptions, similar findings were reported decades ago by Delattre (1963), who observed in a comparative phonetic study that the peak of an ordinary declarative falling contour in American English is aligned relative to the stressed syllable earlier than corresponding peaks in Spanish, French, and German. To be sure, he expressed his observations in terms of the shape of the contour on the stressed syllable ('falling' in American English, 'rising' in the others) rather than as a matter of alignment, but the graphic representations he uses make clear that the differences could equally be described in terms of the alignment of otherwise similar contours.

An analogy to voice onset time (VOT) in stops may make this point clearer. If we say that a language has a voicing contrast in stops, we expect to find a phonetic difference based at least in part on VOT. However, there is no absolute VOT value that corresponds to a given phonological category. For example, we know that VOT can vary from language to language and is subject to a variety of phonological conditioning factors (e.g. velar stops typically have longer VOT than coronal or labial stops) (Cho and Ladefoged 1999). Yet very little of this variation is reflected in phonological description; it makes sense to describe both English and Italian as having a phonological voicing contrast in stops, even though some Italian voiceless stops may be phonetically very similar to some English voiced ones. Voicing is a categorical phonological abstraction; quantitative phonetic details are properly dealt with in a different part of an overall description. Since both VOT and F_0 segment alignment involve the temporal coordination of laryngeal and supralaryngeal gestures, similar ways of describing them in phonology and phonetics seem appropriate. The relation between phonetic detail and phonological representation in the description of segmental anchoring is further discussed by Ladd 2006.

Finally, it should be mentioned in this connection that the perceptual relevance of the articulatory regularities discussed in this subsection is unclear. It certainly seems unlikely that F_0 turning points are directly perceived as such (which is another reason that it probably makes little sense to treat their alignment as a direct reflection of the phonological association of an underlying tone). Among other things, many contours fail to show a clear turning point at an accented syllable, and many turning points are not clear maxima or minima, but merely 'elbows' or inflection points between a relatively level stretch of F_0 and a clearer rise or fall (e.g. D'Imperio and House 1997). It therefore seems more probable that, in some way, the location of the turning point affects the perception of the F_0 level of the accented syllable, both relative to the speaker's range and relative to adjacent syllables. These questions are only just beginning to be investigated empirically (e.g. Dilley 2005; Knight and Nolan 2006; Dilley and Brown 2007; cf. also Chen and Xu 2006) and it is too early to predict what effect this research will have on our understanding.

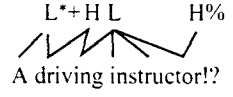
5.1.3 *Compression, truncation, and other effects on alignment*

We have already seen several cases where the details of alignment are conditioned in predictable ways by a variety of different factors. First, there are cases where segmental anchoring of a specific accent type in a specific language is categorically affected by some phonological factor, such as the phonological length of the vowel (Ladd, Mennen, and Schepman 2000; Ladd *et al.* forthcoming)

or the syllable structure (Prieto and Torreira 2007). Second, there are cases where the realisation of a given tonal specification seems to be affected gradiently by phonetic pressure from a closely adjacent tone or boundary; these include the effects investigated by Silverman and Pierrehumbert (1990) and Prieto, van Santen, and Hirschberg (1995), and in more general terms by Xu and Sun (2002). Note that the consistent differences in the alignment of nuclear and prenuclear accents, referred to in section 4.1.1, could be seen as an example either of a categorical phonological or a gradient phonetic effect: we might hypothesise that alignment is directly affected by the phonological status of the accent as nuclear or prenuclear, or, alternatively (as proposed by Hualde 2002), we might attribute the difference between these two to phonetic pressure on nuclear peaks from the upcoming phrase accent.

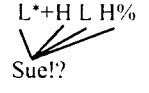
However we interpret alignment effects in nuclear accents, it is clear that they differ from language to language. In the European languages we have been considering, where a nuclear accent consists at a minimum of a pitch accent and an edge tone, the situation frequently arises that two, three, or even more tones are associated with a single syllable in a monosyllabic phrase or utterance, and this creates phonetic pressure that different languages resolve in different ways. Recall the example of the English rise–fall–rise tune applied to two different texts, which was used (section 2.1) to illustrate the usefulness of the AM descriptive framework. In *Sue!?* the tune seems like a single continuous rise–fall–rise, whereas in *a driving instructor!?* we can see that it consists of two clearly distinct parts, the rise–fall at the nuclear syllable and the final boundary rise. In AM terms, the tonal sequence is L*+H_L_H%. In the longer utterance the tones will spread themselves out over the available syllables in something like the following way:

(1)



In the monosyllabic utterance, however, there is only one available syllable, and the pitch therefore rises and falls and rises again:

(2)



The complex pitch movement in (2), in which all the tones are realised on a single syllable, is quite normal in English. (If this were not the case the comparison of *Sue* and *driving instructor* in section 2.1 would have been considerably less convincing!) In some languages, however, there seems to be

a limit to the number of tones that can be realised on a single syllable, the most common limit being two. This can be illustrated with the Hungarian question intonation discussed in section 2.5. As we noted, this is a tonal sequence L*_H_L%, where the H phrase accent is preferentially associated with the penultimate syllable. In two-syllable utterances the phrase accent is squeezed over onto the final syllable, but the distinctive low–rise–fall pattern is present in both cases. In monosyllables, however, the contour is usually reduced to a simple rise. In AM terms, this can be described by saying that no more than two tones can be realised on a single syllable. So, for example, in

(3) sör? 'beer?'

we may say that we have underlyingly

(4) L* H L%
sör

but that only the first two tones of the three-tone question tune are realised with the lone syllable of the utterance. The final L% is left unrealised, and the resulting contour is a rise.

These two patterns of dealing with phonetic pressure in nuclear accents have been referred to (e.g. by Grønnum 1991) as *compression* (for the English type) and *truncation* (for the Hungarian type). Different languages seem to prefer one strategy over the other, though the distinctions are by no means clear-cut. For example, English is, on the whole, a compressing language, but Grabe *et al.* 2000 suggest that there may be regional variation within English. Greek, too, clearly prefers compression, at least for question intonation: when the nuclear accent of the L*_H_L% question tune is applied to a final stressed syllable, the final syllable is lengthened and, in addition, the L* is aligned earlier than with a penultimate or antepenultimate stressed syllable (Arvaniti, Ladd, and Mennen 2006). By contrast, Hungarian and Palermo Italian (Grice 1995a) seem to prefer truncation (Grice's term for truncation is *curtailment*).

Compression and truncation are most conspicuous (and provide the most conspicuous cases of differences between languages) in nuclear accents in absolute final position, but similar effects are found whenever accents are too close together to be realised in full. This topic has been investigated extensively (e.g. Caspers and van Heuven 1993; Xu and Sun 2002; Prieto 2005). Here again, languages appear to have a variety of ways of dealing with phonetic pressure: it seems to be particularly common for F_0 valleys between H peaks to be 'undershot' (i.e. to be realised at a higher F_0 level than they would otherwise be; e.g. Prieto 1998), but it is also possible for one accent type to be replaced by another, to create an accent sequence that can be more easily realised (e.g.

Arvaniti, Ladd, and Mennen 1998; Prieto 2005). The former strategy involves something like compression, while the latter is comparable to truncation.

Compression seems to be clearly a matter of adjusting the phonetic realisation of a contour that remains in some sense unchanged phonologically. Whether truncation involves an actual phonological substitution or merely a phonetic reduction is not always clear. In an example like (4), one could imagine writing a phonological rule to de-associate the truncated tone(s); such tonal rules have often been posited in the autosegmental literature on African tonal systems, and Grice discusses her Italian data in such terms (1995a: 171ff.). However, the line between the two types of description is less clear than it might seem. For example, in Hungarian, it is possible to get all three tones of the question tune realised on a monosyllable if the monosyllable is prolonged for paralinguistic reasons – for example, to express amazement:

(5) L* H L%
|//
söör!? '[you can't mean] beer'

This might suggest that the phonological specification remains unchanged, and the lack of any phonetic fall in pitch at the end of (3) is merely a phonetic adjustment to the fact that the sonorant part of the syllable is not long enough to realise a full rising–falling pitch contour. Either way, though, it is clear that Greek and Hungarian adopt different strategies – whether as a matter of categorical phonological rules or gradient phonetic realisation specifications – for dealing with nuclear accents on monosyllables.

In some languages there is also clear evidence of something like phonological replacement strategies for dealing with this kind of phonetic pressure. In German and Dutch, for example, it appears that speakers generally use a different tune rather than produce a falling–rising pitch movement on a final accented syllable. Suppose someone has left a shopping bag or some money on the counter next to the cash register in a department store and is about to walk away. We might say any of the following, using the high–fall–rise question intonation discussed in 3.2.2:

(6) (a) H* L H%
Ist das IHRE Tüte? 'Is this YOUR bag?'

(b) H* LH%
Ist das Ihre TÜte? 'Is this your BAG?'

(c) H* L H%
Ist das IHR Geld? 'Is this YOUR money?'

Yet, though it is perfectly appropriate pragmatically, it sounds odd phonetically to say

(7) H* LH%
Ist das Ihr GELD? 'Is this your MONEY?'

with the three tones of the high–fall–rise compressed onto the final monosyllable (the same observation is made by Féry 1993: 91). More natural would be a change of tune, to a simple rise:

(8) H* HH%
Ist das Ihr GELD? 'Is this your MONEY?'

Avoidance of high–fall–rise intonation on final stressed syllables is also found in Dutch, which otherwise, like German, uses this tune regularly in questions. Lickley, Schepman, and Ladd (2005: 178f.) showed that in a corpus of Dutch unscripted conversation there were proportionally far fewer high–fall–rise tunes on questions with final accented syllables than would be expected on the basis of cases with penultimate or antepenultimate accented syllables. The 'missing' high–fall–rises seemed to have been replaced by rises, or truncated and realised as falls.

5.1.4 *Phonology of alignment*

Despite the clear evidence that alignment is used as the basis of phonological distinctions and is subject to a great deal of lawful phonetic variation, its theoretical status has remained somewhat unclear. In his thesis (1977), Bruce himself gave no direct phonological representation to the differences of alignment in his notation, treating both Accent 1 and Accent 2 phonologically as sequences of H and L. The alignment distinctions in his analysis resulted from the operation of the phonetic realisation rules, and in a sense were not part of the phonology at all. Given the lexical and morphological significance of the word-accent distinction, this is surely unsatisfactory.

Pierrehumbert (1980), developing Bruce's idea that alignment can be the phonetic basis of phonological distinctions, introduced the star notation (section 3.1.1) as a way of indicating such distinctions in AM descriptions of pitch. Specifically, she proposed that in English there are distinctions between early and late aligned rises, and early and late aligned falls, which can be understood phonologically if we conceive of rises and falls as bitonal pitch accents L+H or H+L. Given such an analysis, the difference between an early aligned accent and a late one is a matter of which of the two tones is associated with the accented syllable. In an early aligned rise, for example, only the H is associated

with the accented syllable, and the L is a 'leading tone'; this is what is expressed by Pierrehumbert's notation L+H*. The star notation was an immediate success, and Bruce himself later reanalysed his own description of Swedish Accent 1 and 2 as a distinction between early aligned H+L* and late aligned H*+L (Bruce 1987).

However, this is scarcely the end of the story. If the star notation were used only to indicate the relative alignment of the two tones in a bitonal accent, it could be seen as equivalent to Grice's metrical notation discussed in section 4.1.3. That is, H+L* and H*+L could be taken as shorthand for the following metrical representations:

(9)

In fact, however, this abstract and purely relational interpretation is at odds with the use of the star to distinguish monotonal pitch accents from non-accentual tones (i.e. the distinction between a pitch accent H* and a phrase accent H-), which suggests that the star is *not* purely relational, but indicates some intrinsic degree of prominence or some other intrinsic phonetic content. This 'phonetic' use of the star notation quickly took hold. As a typical illustration, consider Gibbon's description of the alignment difference between Northern and Southern German: 'In general, Southern dialects are associated with a *right-displaced prominence peak*, that is, the syllable perceived as being accented has low pitch, and a pitch rise, often followed by a peak, occurs on one of the following syllables (ToBI L*+H, similar to Bolinger's Accent C in English (Bolinger 1958))' (Gibbon 1998: 93; emphasis in original). It can be seen that Gibbon is here using L*+H as a phonetic notation for a contour type – an accent with a 'right-displaced prominence peak' – irrespective of any phonological contrast of alignment that may occur in Southern German.

This phonetic use of the star notation has led to considerable inconsistency and theoretical uncertainty. Three issues in particular are important for our discussion here:

(1) Does the star have some intrinsic paradigmatic meaning, distinguishing one type of tone from another (as is suggested by the distinction between monotonal H* and H-)? Or is it essentially a syntagmatic device indicating internal structure within bitonal (or multi-tonal) pitch accents (e.g. signalling which tone is the head or metrically stronger element, as in Grice's notation)?

(2) Even if we follow Pierrehumbert's original understanding that the star indicates which of the two tones is associated with the accented syllable, what detailed phonetic claims are entailed by this association (cf. the distinction between alignment and association discussed above)? How exactly is the starred tone aligned relative to the stressed syllable (e.g. in the middle of the stressed vowel)? What determines the alignment of the leading or trailing tone (e.g. a fixed time interval before or after the starred tone)?

(3) What relationship, if any, is there between the starred tone and the percept of an accent as being 'high' or 'low'? Should L+H* sound high and L*+H sound low?

Much descriptive work in the 1980s and early 1990s implicitly assumed something like the following cluster of answers to these questions: (1) starred tones do have some essential paradigmatic properties, so that the significance of the star is not merely relational; (2) the starred tone is aligned with the nucleus (or perhaps the rhyme) of the accented syllable, and leading or trailing tones are aligned relative to the starred tone; (3) whether the accent is perceived as high or low is relevant to deciding whether the starred tone is H or L.

The inadequacy of these assumptions was discussed extensively by Arvaniti, Ladd, and Mennen (2000) on the basis of their findings of segmental anchoring in Greek (section 5.1.2). One of the questions they had originally set out to answer was whether the starred tone of the prenuclear rising accent is L or H. Even at the beginning of their experiments they were puzzled by the apparent conflict of criteria: the accent sounds high, suggesting that H is the starred tone, but the L, unlike the H, is extremely invariant both in scaling and alignment, suggesting stability of the sort that might be expected to characterise the 'head' tone of the accent. The outcome of their experiments made it clear that most phonetic assumptions about starred tones were simply not supported empirically: neither the L nor the H is in any obvious way aligned with the accented syllable – certainly not with the accented vowel – and nor is one of the two tones aligned a fixed distance from the other. The phonetic facts are relatively clear, but they do not appear to be reducible to a distinction between two notational choices, L*+H and L+H*.

A related point was made by Atterer and Ladd (2004). As we saw in section 5.1.2, they found that prenuclear accentual rises are aligned later in Southern German than in Northern German, and later in Northern German than in English. They suggested that alignment should therefore be seen as a quantitative phonetic variable like voice onset time, and that it is not only misleading, but perhaps even meaningless, to try to categorise rising accents phonetically as being either L+H* or L*+H, and to say that a given language or dialect 'has' L+H* while another 'has' L*+H, just as it is probably pointless to argue about whether English stops in initial /s/-stop clusters are 'really' voiced or 'really' voiceless (cf. the discussion in section 3.2.2). Alignment is a measurable phonetic property of F_0 movements or turning points; categories of alignment are phonological abstractions. The distinction is obscured if the star notation is used as a phonetic shorthand. However, the issue is still not resolved! Non-relational or phonetic use of the star is still common in ToBI systems for various languages, and practical guidelines on ToBI transcription often make reference to whether an accented syllable sounds low or high.

Finally, before leaving the topic of the phonology of alignment, we should also mention briefly the use of so-called secondary association to describe alignment facts. The notion of secondary association was first proposed by Pierrehumbert and Beckman (1988) to describe the behaviour of initial L boundary tones in Tokyo Japanese. Pierrehumbert and Beckman proposed that the underlying association of these tones is with a phonological abstraction (specifically, in their proposal, a node in a prosodic structure, but one could imagine an essentially equivalent analysis expressed in terms of abstract boundaries). In order for such tones to be realised phonetically, they are then secondarily associated with either the first or second mora of the phrase, depending on a variety of conditioning factors, and the two patterns of secondary association are manifested as two distinct patterns of phonetic realisation. This idea of secondary association was subsequently applied by Grice, Ladd, and Arvaniti (2000) to the description of phrase accents in the Eastern European Question Tune. As we saw in section 4.1.4, there are two distinct patterns of phonetic realisation of this tune – one in which the H phrase accent peak occurs on a postnuclear lexically stressed syllable, and one in which it occurs on one of the last two unstressed syllables. Taking their inspiration from Pierrehumbert and Beckman, Grice, Ladd, and Arvaniti suggested that in all cases the primary association is with the boundary, but that, depending on the context, there may be a secondary association either to an unstressed syllable or a postnuclear lexically stressed syllable.

More recently, however, secondary association has been used by various authors to describe details of alignment. The most thorough statement of this idea is a paper by Prieto, D'Imperio, and Gili Fivela (2005). Primarily on the basis of what they analyse as a three-way alignment distinction in rising prenuclear accents in Catalan, they argue that it is necessary to distinguish the primary association of a pitch accent with an accented syllable from the

secondary associations of component tones of the pitch accent with specific points in prosodic structure. These secondary associations can be with structural edges (such as the end of the accented syllable) or with structural constituents (such as the first mora of the accented syllable). It is possible that such phonological analyses of alignment differences will turn out to be justified. Quite independently of Prieto, D'Imperio, and Gili Fivela's work, a recent analysis of Thai by Zsiga and Nitisaroj (2007) proposes a simple account of a variety of acoustic and perceptual data, based on the assumption that the elements of Thai tones are distinctively associated with either the first or the second mora of the syllable, not with the syllable as a whole. However, given the problems discussed earlier (section 5.1.2) with Ladd, Mennen, and Schepman's proposal that the alignment of pitch accents with Dutch long and short vowels could be explained in terms of association of individual tones with syllable edges, arguments for the direct phonological representation of alignment details as distinctions of autosegmental association should probably not be accepted uncritically. The AM approach has led to real progress in our understanding of alignment, but we still have a considerable amount of phonetic research to do.

5.2 Pitch range and the scaling of tonal targets

Having considered the phonetic realisation of tonal targets on the 'horizontal' time dimension, we now turn to the question of the 'vertical' scaling along the dimension of pitch. As we saw in section 2.3, the basic AM approach to intonation involves an analysis in terms of two abstract phonological tone levels and a moderately elaborate set of phonetic realisation rules. Given a string of tones – which is the ultimate representation of the pitch contour in the AM approach – one important task in providing explicit phonetic realisation rules for pitch contours is to establish the F_0 level at which each tone is realised. In one fairly obvious approach, which was developed in some detail for use in experimental speech synthesis systems in the 1980s (e.g. Pierrehumbert 1981; Liberman and Pierrehumbert 1984; Ladd 1987b, 1990b; van den Berg, Gussenhoven, and Rietveld 1992), there are several independent but interacting factors affecting each tone: if the speaker's voice is raised, all tones will be higher than normal; if a tone is downstepped, it will be lower than it would otherwise be; and so on.

Yet pitch is an anomalous phonetic feature. In dealing with segmental speech sounds, we have many points of reference that are essentially valid for all speakers, and there are well-established taxonomic frameworks for both articulatory and acoustic aspects of segmental sounds. There is some degree of inter-speaker variability, but it does not affect the usefulness of descriptions like 'high back unrounded vowel' or 'second formant approximately 1800 Hz'. For pitch, such fixed terms of reference are largely lacking: pitch differs conspicuously from speaker to speaker (e.g. male vs female speech), from occasion to occasion (e.g. bored vs angry speech), and even from one part of an utterance to another (e.g. 'declination' and other similar effects). This appears to mean that we must provide characterisations that are explicitly *relative*.

The question is, relative to what? There are two obvious ways of thinking about the relative nature of pitch features: they may be relative to the speaker's voice, or they may be relative to other parts of an utterance. For reasons that will become clear presently, I will refer to these two approaches as *normalising* and *initialising* respectively. The proper balance between these two points of view has long been a point of contention – or just plain confusion – in theoretical discussions of the phonology of pitch. Before we can discuss the question of how tonal targets are scaled, therefore, we need to deal with the more general issue of pitch range.

5.2.1 *'Suprasegmental' features and the initialising approach*

The need for some relativity in the characterisation of pitch features is the source of the idea that pitch is somehow an intrinsically relational or syntagmatic feature, which can be correctly interpreted only when there is movement or change. This idea, widespread in phonology as well as in phonetics, was explicitly discussed by Jakobson, Fant, and Halle (1952). Their influential scheme of distinctive features posited a fundamental divide between *inherent* features (the vowel and consonant features, which are supposed to be characterisable in absolute acoustic terms) and *prosodic* features (features of pitch, stress, and duration, which depend for their definition on acoustic change or contrast within the context of the utterance). Jakobson, Fant, and Halle claim that any 'opposition of inherent distinctive features [is] definable without any reference to the sequence. No comparison of two points in a time series is involved. Prosodic features, on the other hand, can be defined only with reference to a time series' (p. 13). With specific reference to pitch, they approvingly quote Pike (1948: 18), who says that 'the important feature is the relative height of a syllable in relation to the preceding or following syllables'.

Jakobson, Fant, and Halle's distinction between inherent and prosodic features corresponds almost perfectly with the American structuralist distinction between segmental and suprasegmental phonemes (e.g. Lehiste 1970).

Among other things, the general understanding of pitch features implied by the segmental–suprasegmental (or inherent–prosodic) split is what underlies the 'configurations' position in the 'levels-vs-configurations' debate (cf. section 2.3.1). The 'configurations' point of view describes pitch contours in terms of elements that intrinsically refer to a time series, like 'rise early in the stressed syllable': a rise is a rise, and remains a rise whether it moves from 90 Hz to 120 Hz in a low male voice, or from 200 Hz to 450 Hz in an animated female voice. There is no need to refer to pitch level, and hence no need to deal with range differences anywhere in the description.

More generally, the Jakobson–Fant–Halle phonological theory can best be implemented in what I am calling an 'initialising' description of pitch phonetics. Such a description attempts to provide invariant characterisations of all tones in terms of *what has preceded in the utterance*. (To give an oversimplified illustration, it might state that H tone is realised 6 semitones higher than an immediately preceding L, and half a semitone lower than an immediately preceding H.) The only thing such a model requires in order to derive actual F_0 values is an initial state for each utterance. It does not ever need to refer to characteristics of the speaker's range; all it needs is a starting point, and each successive target can be specified with respect to what has gone before. For example, in such a model, low rise could be distinguished from high rise by the fact that the former starts lower than what precedes and the latter starts higher. This was an explicit part of Bolinger's definitions of English pitch accent types or 'profiles' (Bolinger 1958, 1986), which are all expressed in terms of pitch change to or from the accented syllable.

The initialising approach is clearly capable of modelling many aspects of F_0: for example, it can readily be used to describe local modifications of pitch range for emphasis, changes of topic, and the like. This idea was developed extensively by Crystal (1969: 143–52). In Crystal's transcription system, each syllable bears one of either five or seven possible relations to the immediately preceding syllable: same level, slightly higher, slightly lower, much higher, much lower, and (for stressed syllables only) very much higher, and very much lower. Furthermore, these distinctions can co-occur with what Crystal calls 'step-up' and 'step-down', which affect the overall level of a whole stretch of speech (such as a parenthetical phrase). In addition, each distinctive pitch movement (the nuclear tones) may span a range that is classed as normal, wide, or narrow, according to the distance between its beginning point and its end-point. By combining these dimensions of variation, Crystal is able to transcribe a remarkable amount of phonetic detail without any reference whatsoever to absolute F_0 levels or to the limits of the speaker's range.

There are well-known problems with the Jakobson–Fant–Halle point of view, however, and with the notion of a fundamental split between 'suprasegmental' and 'segmental' phenomena. The most important of these problems is the existence of languages with level tone phonemes, like Yoruba. If we adopt an initialising view, then we will have to define these tone phonemes relative to each other: for example, within some domain to be identified, H is higher than M, which is higher than L (cf. Clements 1979). But this definition cannot always be applied in a given utterance. In many such languages it is quite possible to have utterances that consist of only a single syllable with a single tone, or of a string of syllables all having the same tone. Such cases pose a problem for the formal distinction between inherent and prosodic features: if H can only be defined as locally higher than M or L, then it remains undefined in an utterance with only H tones.[4]

A second phenomenon that suggests the usefulness of defining pitch level relative to the speaker's voice range is the existence of distinctions like 'high rise' versus 'low rise'. Such distinctions are assumed for some tone languages (e.g. Cantonese), and are posited in many intonational descriptions which in theory are based on pitch movement alone. The IPO model of Dutch intonation, for instance, draws a distinction between a Type A Fall, which falls to the bottom of the range, and a Type E Fall, which is a partial or incomplete fall. Many traditional British descriptions of English intonation (beginning with Palmer 1922) distinguish high rise from low rise, and some (notably O'Connor and Arnold 1973) distinguish high fall from low fall. Although, as we just saw, Bolinger and others have tried to define these in syntagmatic terms, we will probably have to acknowledge that, for example, a 'low rise' is low relative to

4 On this problem, Jakobson and Halle suggest the following:

> both alternatives of a prosodic feature co-exist in the code as two terms of an opposition and, moreover, co-occur and produce a contrast within the message. If the message is too brief to include both contrasting units, the feature may be inferred from the substitutive cues offered by the sequence; e.g., ... the register of a monophonemic message [may be inferred] from the modulation span at the onset and/or decay of the vowel (1971: 37)

In fact, to the extent that we understand how pitch perception works in languages with level tones, we know that syntagmatic considerations are involved in some way: for example, in many African languages L tones normally fall somewhat, which helps to distinguish them from M and H tones (see e.g. LaVelle 1974). Nevertheless, Jakobson and Halle's statement seems tantamount to acknowledging that there are phonetic cues to distinctive pitch level even in the absence of syntagmatic contrast, and it is therefore not clear why pitch level is treated as fundamentally different from other phonological distinctions.

the speaker's overall voice range, and that therefore some way of phonetically characterising levels within the speaker's range is ultimately needed.[5]

These and other problems formed the basis of Leben's extensive critique (1973) of the suprasegmental concept. Building on Leben's work, autosegmental phonology has in many respects abandoned the idea of a fundamental distinction between segmental and suprasegmental. One obvious illustration of this is the fact that autosegmental analyses of lexical tone do not assume that the primitives of pitch systems must involve change or movement. If anything, they strive to express all pitch contrasts in terms of static level tones. Even in the description of languages like Chinese, where the motivations for referring to pitch movements are strong (Pike 1948; Wang 1967; Gandour 1978), autosegmental analyses have generally favoured treating contour tones as sequences of level targets like H and L (e.g. Woo 1969; Anderson 1978; Yip 1989; Duanmu 1994; for a review see Yip 2002: ch. 3).

There is also a more fundamental way in which autosegmental and metrical theory have moved away from Jakobson, Fant, and Halle's view. In autosegmental representations, pitch features are exactly like segmental features. This can be seen in the use of the term 'melodic' to refer to both kinds of features, as well as in the analogy between contour tones and diphthongs just mentioned. Stress and duration, on the other hand, are treated in current non-linear phonology in terms of the structural properties of phonological strings, rather than as features or elements of the strings themselves. Traditional suprasegmentals – or the Jakobsonian prosodic features – do not form any sort of natural class within this theoretical picture: if we were to express the autosegmental outlook in Jakobsonian terms, then we would certainly say that pitch is an inherent feature, and that the only prosodic features are stress and duration.

5.2.2 *The normalising approach and the notion of 'tonal space'*

The theoretical considerations just discussed suggest that we should explore the possibilities of a normalising rather than an initialising model of pitch range. A normalising model reifies the notion of pitch range, or what we might call 'tonal space', in terms of some speaker-specific reference points, such as upper and lower F_0 values. Such a model attempts to abstract away from differences between speakers, paralinguistic effects, and so on, and expresses the invariant

5 This issue was aired in the early days of the levels-vs-configurations debate. Sledd (1955) argued that any distinction like low rise versus high rise effectively undermines the levels-vs-configurations dichotomy. However, some British descriptions (notably Schubiger 1958 and especially Crystal 1969) have treated such distinctions as paralinguistic or otherwise orthogonal to the basic distinction between rise and fall, and may thereby be said to maintain a purely 'configurations'-based description.

characterisations of tones in terms of the *idealised speaker tonal space* that results from this process of factoring out sources of variation. For example, instead of stating that H tone is realised some distance higher than L, it might state that H tone is realised at the top of the speaker's current tonal space. The tonal specifications in such a model will make no reference to the phonetic context of the utterance: the beginning of a low-rise configuration will have its lowness defined relative to the idealised tonal space, not – as in an initialising model – to some preceding pitch level.

As we saw in section 2.3.4, the reified 'tonal space' just described is precisely analogous to the reified 'vowel space' that has been part of phonetic practice for nearly a century. Vowel phonemes are defined relative to the vowel space, not in terms of absolute formant values, and not relative to other vowels in an utterance context. We have no trouble when someone says, for example, that vowels have 'more peripheral' realisations when they are stressed, even though for any given speaker (or indeed, for any given vowel) the notion 'more peripheral' will mean something quantitatively different. We do not assume that vowel quality can only be correctly perceived by hearing other vowels in the same utterance (though such syntagmatic effects actually are relevant to perception, as shown in the famous experiment by Ladefoged and Broadbent 1957). The main difference between tonal space and vowel space seems to be the one already mentioned: that tonal space can differ conspicuously from one speaker to another and from one situation to another. This does not seem reason enough to treat pitch phonology as involving an entirely different kind of phonological element.

Probably the simplest possible normalising model of tonal space would specify the highest and lowest values of an individual's overall speaking range, and define the phonetics of F_0 relative to these points. This was the approach taken by Earle (1975) in his study of lexical tones in monosyllabic Vietnamese words. Assigning a value of 100 to the top of each speaker's 'average F_0 range' (i.e. tonal space) and 0 to the bottom, Earle phonetically defines the citation forms of each of the six lexical tone contours in terms of movement along the percentage scale defined for each speaker. Rose (1987), in his study of Wu Chinese, takes a somewhat different approach: his normalised speaker-specific scale is based on a z-score transform[6] rather than on the top

6 A z-score normalisation is based on the statistical distribution of data points – in this case the F_0 values of each individual analysis frame in an utterance or a longer corpus of a speaker's speech. The mean F_0 value is assigned a normalised score of 0, and the size of the units on the normalised scale is determined by the standard deviation: an F_0 value that is 2 standard deviations above the mean has a normalised scale value or z-score of +2.00; an F_0 value that is 1 standard deviation below the mean has a z-score of –1.00; and so on.

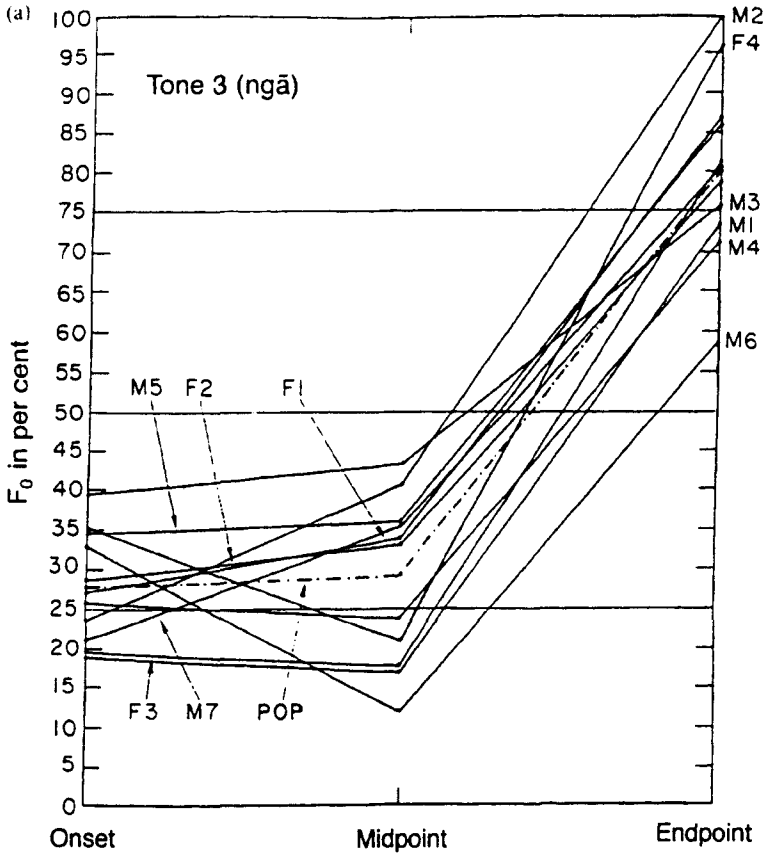


Figure 5.3. *Normalised lexical tone contours. Panel (a), from Earle 1975, shows average starting, middle, and ending points for the Vietnamese Tone 3 (ngā), based on multiple readings by twelve different speakers, plotted on a normalised percentage scale that treats the bottom of each speaker's range as 0 and the top as 100. Panel (b), from Rose 1987, shows pitch traces for Wu Chinese Tones 1 and 3 for seven different speakers, plotted on a normalised z-score scale that treats the speaker's mean F_0 as 0 and defines scale units in terms of the standard deviations around each speaker's mean.*

and the bottom of the range, and indeed he argues that such a normalisation is superior to Earle's. In both cases, however, the conclusion is essentially the same. *Normalised descriptions of the tone contours show a high degree of inter-speaker agreement.* Results from Earle and Rose are shown in figure 5.3.

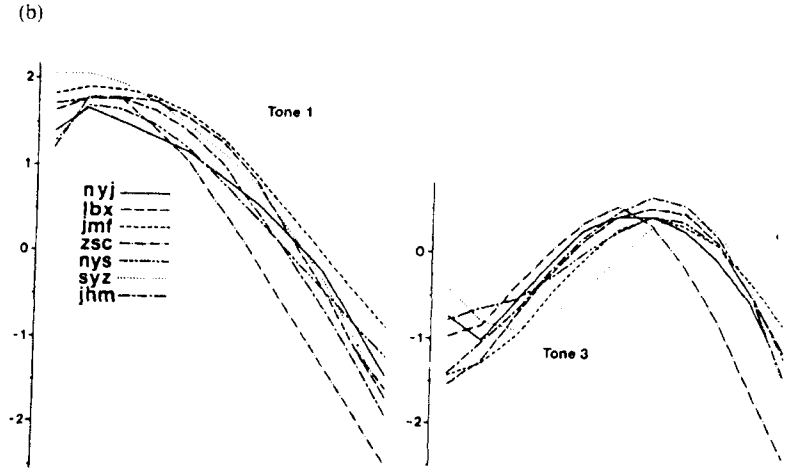


Figure 5.3. *(continued)*

What both normalisations provide is a phonetic description that is not based on the utterance context but can nevertheless yield explicit quantitative predictions of the acoustic characteristics of specific tones spoken by specific speakers in specific contexts. For example, in Earle's definition, Vietnamese Tone 3 (*ngā*) starts at about 28 per cent, rises slightly to about 30 per cent by the midpoint of the syllable, and then rises sharply to about 80 per cent by the end of the syllable. (This is what is shown in figure 5.3.) This compares to Tone 2 (*sắc*), which starts at about 35 per cent, rises to above 50 per cent by the midpoint, and reaches 100 per cent at the end. Both are 'rises' and both span much of the speaking range, but the normalised description makes it possible to predict in detail how they will differ for different speakers.

Comparable findings are found in typologically different tone languages as well. For example, data from Mambila (a Northern Bantoid language of the Cameroon–Nigeria border area), made available to me by Bruce Connell, show the same kind of regularities as those described by Earle and Rose, but in a language with four level tone phonemes. Connell recorded multiple tokens of words with each of the four tone phonemes, and obtained average F_0 values for each tone phoneme for each of five speakers, three males and two females. These figures are given in the upper panel of table 5.1 (for more detail see Connell 2000). Although the absolute values differ quite considerably from one speaker to another, most of the variation between speakers disappears even on a very simple model of the sort suggested by Earle. Specifically, if for each

Table 5.1 *Mean F_0 values for the four lexical tones of Mambila in the speech of five speakers (data supplied by Bruce Connell). Upper panel: mean values in Hz. Lower panel: mean values expressed as a percentage of each speaker's overall pitch range, based on setting the mean High tone value to 100 and the mean Low tone value to 0. For more detail see text.*

	CD♂	SM♂	BJ♂	VM♀	MD♀
High	166.4	120.2	136.0	251.6	199.6
High-Mid	141.9	108.7	120.4	214.0	175.6
Low-Mid	130.3	101.6	115.1	197.0	166.9
Low	116.9	92.2	104.3	171.9	149.1
High	100	100	100	100	100
High-Mid	51	59	51	53	52
Low-Mid	27	34	34	31	35
Low	0	0	0	0	0

speaker we set their mean Low tone F_0 value to 0 and their mean High tone F_0 value to 100, and express the mean values of the High-Mid and Low-Mid tones as percentage values on those speaker-specific normalised scales, the empirical values of the two Mid tones are similar for all speakers. This can be seen in the lower panel of table 5.1. The fact the we can apply such a model to level tones as easily as to dynamic tones strongly suggests that it is both possible and useful to talk of some sort of phonetic equivalence of pitch levels across speakers, and that therefore the notion of 'tonal space' is worth developing.

Whatever the eventual form of a phonetic model of tonal space, one important fact emerges clearly from the instrumental data just summarised: linguistic equivalence of pitch between speakers with different ranges is not based on anything like a musical scale. In a musical scale, equivalent intervals in different ranges are based on *equivalent frequency ratios*. For example, the interval 'perfect fifth' always involves a frequency ratio of 3:2; if one note has a frequency of 3x Hz and another has a frequency of 2x Hz, the two notes will always sound a fifth apart. When men and women sing 'in unison' together, they normally sing in different octaves – the women an octave higher than the men. This means that the frequency of the women's notes is always twice the frequency of the men's notes, but the frequency ratios are the same in both voices. All known musical interval systems are based on such constant ratios (see Patel 2007 for a good survey of this and related issues).

In language, however, there is no such constancy. In the Mambila data in table 5.1, for example, the frequency ratio between High tone and High-Mid tone varies from 1.175 (2.8 semitones) for speaker CD to 1.105 (1.7 semitones) for speaker SM. Nevertheless, these intervals count as linguistically equivalent, in a way that makes reference to the range of frequencies normally used by a given speaker. For both CD and SM, that is, the High-Mid tone is scaled just above the middle of the tonal space: the figures 51 per cent and 59 per cent in table 5.1 indicate that the interval between High and High-Mid is slightly less than half the speaker-specific interval between High and Low. Listeners thus can – indeed must – extract at least two kinds of information from pitch level regularities of the sort shown in table 5.1: "linguistically" they can distinguish High from High-Mid, and at the same time they can assess "paralinguistically" that SM has a narrower tonal space than CD. There is no obvious analogue for this kind of dual analysis in music.

It is therefore important to be aware from the outset that pitch range or tone space is not a single phonetic variable. Broadly speaking, we need to allow for at least two partially independent dimensions on which individual F_0 ranges can differ: differences of overall level, and differences in *span* (the range of frequencies used). One speaker's voice may be higher than another, but in addition one speaker's habitual frequency span may be wider or narrower than another's. For example, in table 5.1 we can see that female speaker VM's voice range is unambiguously higher than male speaker SM's, because there is no overlap at all between them; but we can also see that male speaker CD's voice range, though only slightly higher than that of male speaker BJ, is distinctly *wider* than BJ's (50 Hz or 6.1 semitones for CD, compared to 32 Hz or 4.6 semitones for BJ). In principle, I believe it should also be possible to find cases like the hypothetical male data illustrated in table 5.2 (see also figure 5.4 in section 5.2.3 below, and Liberman 2007), in which two speakers have voices such that one speaker's tonal space is completely included in another's, although we do not find such a case in Connell's five-speaker data set. In any case, it is clear that level and span are at least partially independent.

This partial independence means that we cannot normalise away from inter-speaker pitch differences merely by transforming F_0 data from the linear Hz scale to some other scale, such as the logarithmic semitone scale, or the 'Equivalent Rectangular Bandwidth' (ERB) scale proposed on the basis of psychophysical research by Glasberg and Moore (1990). Hermes and his coworkers (Hermes and van Gestel 1991; Hermes and Rump 1994) have shown very clearly that for certain kinds of simple pitch movements resynthesised in different overall ranges from the same voice source, the perceptual equality of

Table 5.2 *Hypothetical data based on the data in table 5.1. Speaker A's low tone is lower than Speaker B's, and his high tone is higher, but the same proportions are maintained in the way both speakers scale the two mid tones. See further figure 5.4 below.*

	A	B
High	160	125
High-Mid	124	112
Low-Mid	104	104
Low	80	95
High	100	100
High-Mid	55	56
Low-Mid	30	30
Low	0	0

the size of the pitch movement is best expressed on an ERB scale, which is approximately midway between the Hz scale and the semitone scale. However, this conclusion depends on the specific experimental tasks involved, and may depend on the fact that the stimuli come from a single voice. It is clear that for multi-speaker data of the sort presented in table 5.1, the ERB scale will be no more successful than the semitone scale at eliminating differences of span.

To be sure, level and span do tend to co-vary: the higher the level, the wider the span *if measured in Hz*. Women typically have wider span than men if span is measured in Hz, though the differences are considerably reduced (or in some cases even reversed) if span is measured in ERB or semitones. This covariance is certainly one of the reasons why level and span are often not clearly distinguished, but are lumped together under the catch-all term 'pitch range'. Moreover, paralinguistic modifications of level ('raising one's voice') may be difficult to distinguish from modifications of span: in a language like English, where most accents are marked by pitch peaks (or H tones), raising and widening will be difficult to distinguish empirically, since the effect of both is to raise high tones. Nevertheless, as the data presented so far suggest, and as we shall see in more detail in the next section, most observable differences of span cannot simply be mathematically transformed out of existence by the choice of a specific measurement unit. Whatever their differences in F_0 level, some speakers use a fairly wide span and others a comparatively narrow one, and any pitch range model must therefore treat the two separately.

5.2.3 *Experimental studies of pitch range and scaling*

It might be thought that monosyllabic citation forms in tone languages are probably the likeliest place to find regularities of the sort just described. Yet numerous studies of longer utterances, both in other tone languages, and in several non-tonal European languages, point to the same conclusion: the scaling of pitch targets is highly systematic, both within and across speakers. Before considering this topic in more detail, it will be useful to outline in a general way what such studies are looking for, and in particular to sketch what we might expect to find if the 'tonal space' notion can be generalised beyond lexical tone languages to pitch scaling in intonation.

As we saw, Earle compared the scaling of three points in each tone contour – three targets, to use the terminology of chapters 2 and 3. He found that, by defining a speaker-specific percentage scale, he could give a quantitative definition of the scaling of the targets that was approximately valid for all speakers. If this kind of regularity is not confined to tone languages, then we would expect to find similar consistencies in the scaling of intonational targets across speakers in utterances in the European languages. For example, consider the sentence

(10) 

I've been there before.

spoken with the neutral declarative intonation shown. We might identify four targets, an initial L, an accentual H* on *been* and an immediately following L that creates the 'elbow' on *there*, and a final L%. Hypothetical non-normalised data for these four targets as spoken by three speakers are shown in figure 5.4. It can be seen that there is one male speaker and two female speakers, one of whom has a wide span and the other of whom has a narrow span. However, if we normalise the three individual pitch ranges so that the H* represents 100 per cent and the L% represents 0 per cent, then for all three speakers the four target points come out the same, as shown in figure 5.5. This is the kind of finding that would support a normalising model.

With this introduction, let us now turn to some experimental results. First, let us look at studies of the same contour spoken by different speakers. If certain targets are scaled by all speakers at the same level on their own speaker-specific scale, then we would expect to find a high correlation between the target values of one speaker and those of another. This expectation is borne out by data from several studies, including Thorsen 1985, 1986; Liberman and Pierrehumbert 1984; Ladd and Terken 1995; and Shriberg *et al.* 1996.

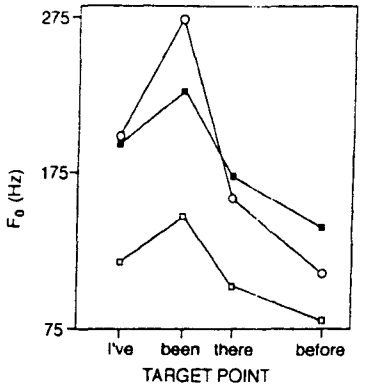


Figure 5.4. *Hypothetical average data for four target points in three speakers' pronunciation of the English declarative sentence* I've been there before. *The contours are impressionistically similar, despite differences of level and span.*

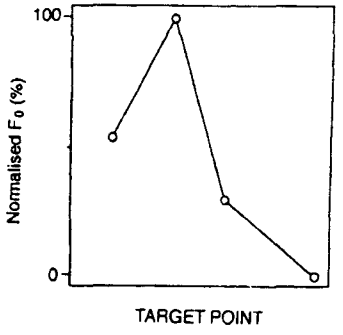


Figure 5.5. *The hypothetical data from figure 5.4, replotted on a normalised percentage scale like that used by Earle 1975. The normalisation abstracts away from the differences of level and span seen in figure 5.4.*

Here I consider in detail data from two studies of Standard Danish by Nina Grønnum (published as Thorsen 1985, 1986; I am grateful to Nina Grønnum for making the raw data available to me). In both studies Grønnum recorded six readings by four different speakers of eight different sentences. The sentences range in length from one pitch accent to eight pitch accents. For each sentence, I have taken the targets to be the beginning and ending F_0 values, and the valley and peak associated with each pitch accent. For each speaker I have thus

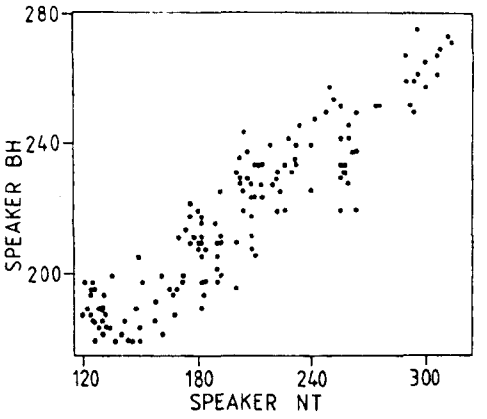


Figure 5.6. *Correlation between target F_0 values in two speakers' readings of the same sentences in Grønnum's studies of Danish sentence intonation (Thorsen 1985, 1986). Each value plotted is the mean of six repetitions. For more detail see text.*

considered 160 targets (2 studies × (2 targets per accent for (1+2+3+4+5+6+7+8) accents+a beginning and an end target for each of 8 sentences)). For each target I have taken the mean of a speaker's six repetitions to represent the intended F_0 level for that target for that speaker.

Figure 5.6 plots all the average target values for one speaker against those for another speaker. In a general way, what this plot shows is that a target that is very low in one speaker's range will be very low in another's; a target that is moderately high for one will be moderately high for another; and so on. The correlation between the target values for the two speakers shown in figure 5.6 is an extremely high 0.92. For all pairs of speakers in Grønnum's two studies, the correlations are never below 0.89. The same overall picture has been found in three studies of other languages in which I have been involved as collaborator or supervisor (Ladd and Terken 1995 and Shriberg, *et al.* 1996 on fifteen speakers of Standard Dutch; Patterson and Ladd 1999 and Patterson 2000 on sixteen speakers of Standard Southern British English ('RP') and sixteen speakers of Scottish Standard English; Yuen 2003 on three speakers of Cantonese). Unfortunately, detailed data from these studies have never been published in full; the best documented of these studies was the one on Dutch. However, none of them provides any reason to think that Grønnum's results are atypical.

Similar regularities to those illustrated in figure 5.6 can be seen when we look at paralinguistic within-speaker variation in pitch range – that is, the raising and lowering of the voice by an individual speaker. The use of experimentally induced modifications of overall pitch range as a means of studying the systematic relations between tonal targets was pioneered by Liberman and Pierrehumbert (Pierrehumbert 1980; Liberman and Pierrehumbert 1984), and the technique has since been used in several other studies, including Bruce 1982; Pierrehumbert and Beckman 1988; Liberman *et al.* 1993; Ladd and Terken 1995; and Shriberg *et al.*1996. (Note that Liberman *et al.* 1993 deal with Igbo, a tone language.) If we assume that inter-speaker differences and within-speaker paralinguistic modification are comparable,[7] we would expect them to be similar in their quantitative manifestations. In particular, we would expect a speaker's target scaling in a normal speaking voice to correlate with his or her own target scaling with a raised or lowered voice.

This expectation is borne out by all the studies just cited. If we calculate correlations between different pitch ranges in the same way as we calculated correlations between speakers in Grønnum's data, we get similar results. Figure 5.7 is based on a study by Bruce (1982), in which a speaker produced the same contours in a low range (as if 'detached') and a high range (as if 'involved'). It shows the two sets of target points – from the detached and involved readings – plotted against each other. Once again the correlation is about 0.90. Similar results based on the much larger Dutch data set mentioned above are reported by Ladd and Terken 1995 and Shriberg *et al.* 1996.

5.2.4 *Modelling pitch range*

While all this evidence indicates in a preliminary way that it may be possible to define speaker-specific scales for use in a normalising model of the phonetics of pitch, quite a number of problems – both theoretical and empirical – remain. The most obvious is that correlation coefficients are too crude to reveal a number of fairly clear quantitative properties of any speaker-specific scale, or

7 Though the two kinds of pitch range differences are comparable in theory, they raise somewhat different methodological problems, because in order to get 'the same' contour spoken in different ranges, we have to instruct speakers to deliberately raise or lower their voices – with uncertain effects on the naturalness of the resulting speech. Also, if we instruct speakers to sound angry or surprised or bored or whatever, we may get a variety of other effects (on speech rate, on voice quality, and possibly on the tune itself), so that we cannot be sure that we are dealing with 'the same' contour. Despite these difficulties, however, results from the studies cited in the text suggest that they have succeeded in getting speakers to produce the same contours in different pitch ranges.

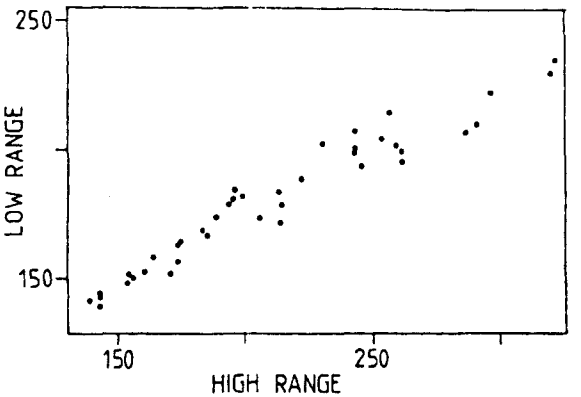


Figure 5.7. *Correlation between target F_0 values in one speaker's 'involved' and 'detached' readings of the same sentences, based on data from Bruce 1982. Each value plotted is the mean of six repetitions.*

to express a number of fairly clear quantitative regularities in studies of pitch range. In order to capture these regularities, more sophisticated models of pitch range are required.

The most obvious quantitative regularity not captured by a correlation coefficient is the fact that the bottom of the speaking range is a fairly constant feature of an individual's voice. As we saw in section 2.3.3.1, numerous studies of both read and spontaneous speech, in a variety of languages (including lexical tone languages), have shown that an utterance-final fall in pitch reaches a low F_0 level that is largely unaffected by raising or lowering the voice or by other within-speaker range changes. What this means is that raising the voice, broadly speaking, involves *expanding the pitch span from the bottom up*.

This being the case, the simplest quantitative model of within-speaker pitch range modifications would take the utterance-final low as a speaker-specific zero level or reference frequency (F_R); express normal target values as a function of F_R; and then scale all target values up or down by a constant factor when the voice is raised or lowered. For example, suppose we determined that in normal speaking voice the average contour onset F_0 value is 1.5 times the value of F_R (i.e. 7 semitones above F_R), while the value of an average high peak is twice as high as F_R (i.e. 12 semitones, or one octave, higher). We then assign those targets abstract pitch values of 1.5 and 2, respectively. If raising the voice increases the onset value to $1.75(F_R)$ – that is, by a factor of 1.167 – the model would then predict that the peak value would be increased proportionally, to $2.33(F_R)$. This kind of model is shown graphically in figure 5.8.

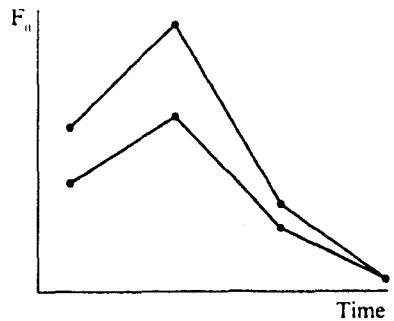


Figure 5.8. *Hypothetical data predicted by a simple model of pitch range modification within a single speaker. The bottom of the range remains constant, and all other target values (filled circles) are scaled up or down when the voice is raised (upper line) or lowered (lower line), in such a way as to maintain constant proportions between target values.*

Mathematically, the model can be expressed as

(11) $F_0 = F_R \cdot T \cdot r$

where F_R is the Hz value of the bottom-of-the-range reference level, T is the abstract pitch value for any given target in normal range, and r is a range multiplier whose value is 1.00 for normal range. This model could readily be extended to the description of range differences between speakers as well, in something like the following:

(12) $F_0 = F_R \cdot T \cdot N \cdot r$

In this, we assign an *invariant* abstract pitch value T to each target, and the factor N normalises across speakers by adjusting the T values for each speaker's span. Thus, building on the example just given, we might treat 1.5 and 2 as invariant abstract pitch values for contour onset and typical high peak. A speaker whose actual onset and peak values were 1.5 and 2 times his or her F_R value would be said to have an N of 1.00; a speaker with a narrow span whose actual onset and peak values were on $1.125(F_R)$ and $1.5(F_R)$ respectively would be assigned an N of 0.75; a speaker with a wide span and an N of (say) 1.15 would be predicted to have onset and peak values of $1.725(F_R)$ and $2.3(F_R)$ respectively.

There are, however, at least three reasons why such a model is inadequate. First, the apparent constancy of the utterance-final low has been called into question on at least two counts: Ladd and Terken (1995) and Shriberg *et al.* (1996) showed that for most speakers, the utterance-final low is affected slightly by extreme raising of the voice; Hirschberg and Pierrehumbert (1986) suggest that the constancy of the utterance-final low is at least in part an artifact of measurement difficulties, and state that manipulating the precise value of the utterance-final low to reflect discourse structure gives better-sounding synthetic speech. Second, a number of experiments by Gussenhoven and his colleagues (summarised in Gussenhoven *et al.* 1997) have shown that manipulating the utterance-final low in a stimulus sentence has no effect whatever on the perceived prominence of accent peaks; this suggests that the notion of 'reference frequency' or 'bottom of speaker-specific scale' cannot be identified in any simple way with the actual utterance-final low in a given utterance. And third, even if there were not such objections, a simple multiplicative model of the sort shown in (12) *does not work*, in the very basic sense that it fails to make accurate predictions about the quantitative regularities that have been observed in range-modification and range-comparison experiments of the sort described above. A more complicated quantitative model is required.

Various attempts at such a model have been made – for example by Ladd (1987b, 1990b); Clements (1990); Traunmüller and Eriksson (1994); and Patterson (2000) – but the most extensive line of work in this area is undoubtedly that of Liberman, Pierrehumbert, and their coworkers (Pierrehumbert 1980; Liberman and Pierrehumbert 1984; Pierrehumbert and Beckman 1988; Liberman *et al.* 1993). The details are different in each case, but the basic idea in all of them, as expressed by Liberman *et al.* 1993, is that pitch range modification involves some *additive* component in addition to the multiplicative factors in (12). Informally speaking, what this means is that raising the voice is *not* just a matter of scaling targets up proportionally from a fixed reference level at the bottom of the range, but involves some modification of the fixed reference level as well. One possible pattern of pitch range modification along these lines is shown in figure 5.9. There are a variety of ways in which this informal idea can be expressed mathematically; unfortunately, present data are insufficient to distinguish between them, and it therefore seems appropriate to omit further mathematical discussion from this brief review.

5.2.5 *Pitch range in AM intonational descriptions*

The preceding sections suggest that an empirically adequate normalising model of pitch range is a real possibility. In this section we turn to the more concrete question of how such a model can be incorporated into AM descriptions of tonal phonology. In particular, we discuss the distinction between what we might call *intrinsic* and *extrinsic* factors in the scaling of tonal targets. By intrinsic factors I mean those that involve the relative position of tonal targets within the 'tonal

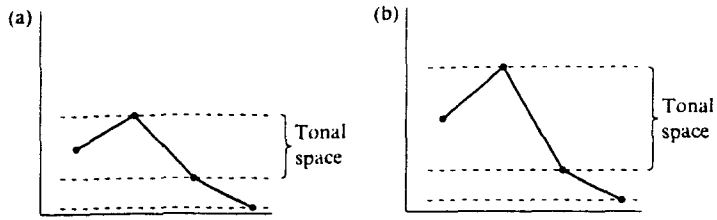


Figure 5.9. *One possible model of pitch range modification, in which the width of the tonal space is expanded or contracted partly independently of the height of the tonal space above the bottom of the range. This independent variability means that the proportions between targets do not all remain constant; for example, the 'elbow' in the contour is proportionally closer to the bottom of the range in (b) raised voice than in (a) normal or lowered voice.*

space', while I intend extrinsic factors to refer to modifications of the tonal space itself. In terms of the kind of quantitative model given in (12) above, intrinsic factors are expressed by the term T of the equation, while extrinsic factors are the speaker variables for level and span (F_R and N), and the utterance variable for overall range (r). If we want our linguistic description of intonation to map onto a normalising model of pitch range, then we must know which phenomena are to be modelled in which way.

Some things seem uncontroversial. A clear example of an intrinsic difference would be that between an H tone and an L tone. Other things being equal, H is higher than L, intrinsically, by the very nature of the two tones in a language that has both. Differences of the sort documented by Earle and Rose in their tone normalisation studies are also intrinsic: the endpoint of a Vietnamese Tone 2 (*sắc*) is – again, 'other things being equal' – higher than the endpoint of a Tone 3 (*ngã*). As for extrinsic factors, it seems obvious that they will include differences between different speakers' overall speaking ranges. The actual F_0 values corresponding to H tone and L tone obviously depend on whether they are spoken by a man or a woman, by a person with a monotonous voice or a person with a lively voice. The acoustic realisation of the tonal space depends crucially on the speaker and the paralinguistic context, but neither the position of the tones within the tonal space nor their linguistic identity is thereby affected.

However, many scaling effects cannot be so neatly pigeonholed. The clearest example of an intermediate phenomenon is downstep, particularly the kind of 'non-automatic' downstep seen in some African languages. In two-syllable

words in Efik, for example, after H tone on the first syllable there is a three-way distinction between H, !H (downstepped H), and L, as in the following example from Connell and Ladd 1990:

(13)
(a) H H
ọbọñ 'mosquito'
(b) H !H
ọbọñ 'chief'
(c) H L
ọbọñ 'cane'

Recognition that the second syllable in (13b) is a downstepped H and not an M dates only from Winston 1960, and seems to have been a hard-won insight. The evidence for the downstep analysis is compelling: downstep 'happens' to an identifiable tone at an identifiable point in the utterance, but the affected tone establishes a new ceiling for following tones, so that a subsequent tone at the same level is unambiguously H rather than M. This strongly suggests that downstep is an 'extrinsic' scaling factor – the scaling of the tone in question does not involve the specification of a particular level within the tonal space, but the modification of the tonal space itself. Yet at the same time it seems intuitively odd to lump such a clearly linguistic phenomenon as downstep with paralinguistic and extralinguistic factors like speaker identity. As a result, there is still not much agreement on how to treat these hard-to-classify factors in AM intonational phonology.

One of the key issues is whether a given distinction should be represented in the tonal string. Pierrehumbert's analysis of English provides examples of different answers to this question. Some distinctions are clearly treated as part of the tonal string, and hence potentially to be modelled using intrinsic scaling specifications; others are consistently ignored in the tonal string, and therefore clearly treated as involving extrinsic pitch range effects; still others are treated in different ways in different versions of the analysis.

First consider the effects that are consistently omitted from the tonal string. This category includes expansions and contractions of pitch range due to local emphasis and relative prominence generally, a topic discussed briefly by Pierrehumbert (1980: ch. 1), but otherwise left largely unexplored. More interestingly, this category also includes the constant quantitative relations between the accent peaks in 'answer–background' and 'background–answer' sentences discovered by Liberman and Pierrehumbert in the *Anna/Manny* experiment (see section 2.3.3). The phonological analysis of the accents in these contours is the same (H* in the original description of Pierrehumbert's 1980,

unspecified in Liberman and Pierrehumbert 1984; cf. section 4.2.2), and the 'answer/background' relation is specified separately. The nature of this separate specification is never made very clear – in that respect it is rather like the specification of alignment in Bruce's original analysis of the Swedish word-accent distinction (cf. section 5.1.4) – but it is certainly not phonological. In fact, Beckman and Pierrehumbert (1992) seem to suggest that it is not any sort of linguistic specification at all, but merely a consistent way of controlling paralinguistic pitch range that speakers adopt in response to a particular experimental task.

The treatment of these effects can be compared with the treatment of downstep. Downstep is clearly phonological in Pierrehumbert's model, always represented in the tonal string. As we have seen several times already, Pierrehumbert treats downstep as the result of a phonetic realisation rule that operates on certain specific strings of tones and not on others, though the actual model of phonetic realisation seems to involve both intrinsic and extrinsic factors. Certain aspects of downstep are modelled as 'extrinsic' modifications of something like the tonal space (see especially the model in Pierrehumbert and Beckman 1988; ch. 7). But in the case of the H+L* accent,[8] the scaling specifications are clearly intrinsic: as we saw in section 3.1.3, in Pierrehumbert's analysis this has a *low* (L*) tone scaled at what might otherwise be considered to be the level of a downstepped *high*, that is, the model provides different scaling specifications within an unmodified tonal space.

Finally, at least one English distinction has been treated in two different ways in Pierrehumbert's work, namely the distinction that she originally analysed as involving the H*+H accent. In the original analysis, H*+H was used to represent a high accent that is followed by a sustained high level transition to a following high accent, in contrast with the H* accent, which is followed by a 'sagging' transition when the next accent is high (cf. figure 3.2). In this analysis, the H trailing tone of the H*+H was said to spread to the following syllables, yielding the characteristic sustained high pitch. However, the H*+H accent had certain theoretically undesirable properties that led Beckman and Pierrehumbert to revise the analysis to exclude the H*+H altogether from the inventory of English accent types. They state (1986: 306) that 'we would now analyse [the sustained transition between H* accents] as involving ordinary H* accents produced in an elevated but compressed pitch range'. In other words, the

8 This refers to H+L* in Pierrehumbert's original sense, i.e. ToBI H+!H* (cf. section 3.1.3, and chapter 4 footnote 1).

phonetic details that originally led Pierrehumbert to posit the H*+H accent are reanalysed as extrinsic scaling factors affecting the sagging transition between H* accents.

This abrupt change reveals the extent to which the AM theory – like every other approach to intonation with which I am familiar – has a problem with pitch range. No real justification is given for the reanalysis: Beckman and Pierrehumbert state that it is 'a natural outcome of the new treatment of pitch range introduced by Liberman and Pierrehumbert (1984)', but the H*+H is never mentioned in Liberman and Pierrehumbert, and there is no elaboration on the claim that this is a 'natural outcome' of their work. No detail is ever provided about when the pitch range can be 'elevated and compressed'.[9] No other intonational distinctions are analysed in exactly such terms, although, as I noted above, Liberman and Pierrehumbert do introduce the idea of extrinsic scaling factors in their analysis of the distinction involved in the *Anna/Manny* experiment. The theoretically embarrassing H*+H is thus done away with, but the more general lack of theoretical clarity about pitch range remains unaddressed.

This issue is taken up by Sosa (1999) in his extensive AM description of Spanish intonation. Sosa rejects Beckman and Pierrehumbert's reanalysis of the English H*+H. Adopting a strongly 'initialising' approach to the description of tonal scaling, he argues for *extending* the type of approach involved in the original H*+H analysis, and for representing *more* pitch range distinctions directly in the tonal string. His inventory of Spanish accent types includes, among others: H*+L, to trigger downstep in a following accent, as in Pierrehumbert's analysis of English; H*+H, which represents the same sort of accent type as in Pierrehumbert's original analysis of English; and H+H*, which indicates extra height on a high nuclear accent in a question. Of the H+H* he says:

> The reason we consider this pitch accent necessary is phonetic: this pitch accent is responsible for raising the tone of this syllable to higher levels than a simple H* tone could do. It would not take account of the melodic facts to state, for example, that the suspension of declination is responsible for the horizontal contour of the [prenuclear contour], and that then the H* pitch accent takes care of the last raising of pitch on the nucleus, since ... there are differences between Caracas dialect and other dialects which have

9 As we just saw in section 5.2.4, when pitch range is raised it normally causes the tonal space to widen, not narrow. Thus the idea of an 'elevated but compressed' pitch range, though obviously not impossible a priori, may be difficult to reconcile with most findings on pitch range modification.

> suspension of declination but do not have a notable raising of pitch in relation to the relative height of the body of the utterance ... Thus, our positing of the H+H* pitch accent takes account of the phonetic details of [yes–no] questions in this dialect, and in addition it has the implication that the information that yields the rise is present underlyingly, and is different from what happens in other dialects. (1999: 157; my translation)

Sosa's guiding principle might be stated as follows: if an intonational distinction is phonological, then it must be represented in terms of different tonal strings. Except for clearly extralinguistic effects such as inter-speaker differences of range, there is no place in Sosa's analysis for extrinsic scaling factors anywhere. Even the higher overall pitch of yes–no questions is analysed by Sosa as involving a difference of tonal sequence, though such a phenomenon is reported or assumed for many languages and might arguably be treated as paralinguistic and hence as an extrinsic scaling factor. The overall raising in questions, according to Sosa, reflects the occurrence of an initial high boundary tone (H%), restricted to yes–no questions, which triggers an initial upstep. This initial upstep 'increases the frequency of the first accented syllable ... to a level noticeably higher than the normal level of the utterance. On the basis of this initial height, the phonetic implementation rules assign numerical values in Hz to the later tones, which results in the effect of greater pitch height for the whole interrogative [prenuclear contour] utterance up until the [nucleus]' (1999: 152; my translation).

As will be clear from the discussion in the previous section, I do not agree with Sosa's approach. I think there is an important role in the description of intonation for extrinsic scaling factors operating at well-defined points in the tonal string; among other things, I think this will simplify the tonal representations of languages like Spanish or English. But Sosa's work is relevant to the discussion here in two important ways. First, it has the virtue of making its theoretical criteria about pitch range explicit. It proceeds from a kind of analytical null hypothesis, which is that intonational distinctions are assumed to be phonological *and represented in the tonal string* unless there is a clear basis for treating them otherwise. Second, Sosa's work shows that the proper treatment of pitch range phenomena in AM intonational phonology is by no means settled. Pierrehumbert's solution – to treat downstep as a quasi-intrinsic effect represented in the tonal string, and to treat all other scaling effects as extrinsic and essentially non-phonological – is not the only reasonable possibility within the general framework of AM assumptions. We return to consider another possibility in chapter 8.

III Phrasing and prominence

6 *Patterns of sentence stress*

At the beginning of the book I proposed that the phenomena of intonation can be understood as involving two essentially orthogonal dimensions, which I referred to as 'pitch' and 'prominence/phrasing'. In the preceding three chapters we have considered in some detail various aspects of the description of pitch patterns. In this and the remaining two chapters we now go on to discuss the treatment of prominence and phrasing in an AM theory of intonation. Roughly speaking, in this chapter we consider the relations between sentence-level prominence (or sentence stress) and focus – which words in a sentence are prominent and why; and in the next we consider the specifically phonological aspects of sentence-level prominence. In the final chapter we discuss prosodic phrasing. By the end of the book I hope to have motivated the basic division of intonation into pitch and prominence/phrasing, and in particular to have shown why it is appropriate to regard 'prominence' as just one of several phenomena that can be accounted for by an adequate 'metrical' theory of prosodic structure.

6.1 Sentence stress and broad focus

It is now generally accepted that the pattern of sentence stress in an utterance reflects the utterance's intended focus, but there is a good deal of disagreement and confusion about just how it does this, and about what 'focus' actually involves. Much of the disagreement and confusion is about issues in syntax and semantics, not phonology, which makes it difficult to do justice to the whole question without straying fairly far from the announced topic of the book. However, the key phenomenon, which I will refer to here as *broad focus*, has clear implications for intonational phonology. In order to be able discuss the phonological questions without treating the syntactic and semantic issues

too cavalierly, I will introduce the general problem of broad focus informally through a series of simple examples.[1]

6.1.1 'Normal stress' or 'highlighting'?

Consider the phrase *five francs*. This could be used in a specific context where the number of francs was at issue, in which case we may say that the focus is on *five*:

(1) I didn't give him three francs, I gave him five francs.

In this case, there would normally be a pitch accent on *five* and none on *francs*, and thus a direct relation between accent and focus. A simple notation for this pattern is *FIVE francs*. The same phrase could also be used in a context where the unit of currency was at issue, in which case we may say that the focus is on *francs*:

(2) I didn't give him five pounds, I gave him five francs.

In this case there would often be a prominent accent on *francs*, and either no accent on *five* or only a weak one. Once again, there is a direct relation between accent and focus, and we can notate this pattern informally as *five FRANCS*. (As we shall see in chapter 7, there is an important question about what happens phonetically on *five* in this case.)

Finally, in relatively unusual circumstances, there is also what we might call a 'double-focus' pattern, with very prominent accents on both *five* and *francs*. This might be found in very deliberate speech in a context like (3):

(3) I didn't give him SEven EUros, I gave him FIVE FRANCS.

The first clause of this example – *I didn't give him seven euros* – also illustrates a basic feature of pitch accents, namely that when a pitch accent occurs on a word containing more than one syllable, the accent occurs on the lexically stressed syllable of the word (*se-* of *seven* and *eu-* of *euros*).

In all these cases there is a direct relation between accent and focus: individual words are highlighted both phonetically and pragmatically. However, sentence stress cannot always be explained so straightforwardly. In addition to the three fairly specific contexts just sketched, it is also possible for the phrase *five francs*

1 As noted by Beaver *et al.* (2007), there is an increasing divide between those who study the syntactic/semantic issues surrounding focus and those concerned with phonological and phonetic questions. My discussion here may strike some as symptomatic of this state of affairs.

to be used in a wide variety of other contexts in which the focus is not on either word alone but, as it were, on the whole phrase:

(4) I didn't give him { a dollar / fifty centimes / my notebook / your camera / the car keys / a sandwich / a lot of money }. I gave him five francs.

This is clearly distinct in meaning or function from any of the first three examples. In those cases, one or both of the words in the phrase *five francs* is contrasted to other possible words from fairly specific sets of numbers and currency names; here, the phrase *five francs* is contrasted *as a unit* to some other phrase from a more or less unlimited set of possibilities. This difference appears to justify distinguishing 'narrow focus' on a single word, as in (1) or (2), from 'broad focus' on a larger constituent, as in (4).

Issues surrounding broad focus have been at the centre of debates about the description of sentence stress for well over half a century. The problem is as follows. Despite the clear meaning difference, the sentence stress pattern that signals broad focus on the larger constituent in (4) is similar or identical to the pattern that signals narrow focus on the single word *francs* in (2). That is, sentence stress involves an asymmetry: when the main accent is on *five*, the intended meaning is narrow focus on *five*, but when the main accent is on *francs*, the intended meaning can be either narrow focus on *francs* or broad focus on the whole phrase. If the sentence is intended to convey focus on a whole phrase or constituent, on what basis is a single word selected to bear the main accent?[2]

In the 1950s and 1960s the lines were drawn between two distinct and competing approaches to broad focus, which we might call the normal stress

2 The terms broad and narrow focus were suggested in Ladd 1980a: ch. 4, but were not used in exactly the way just illustrated. Originally, 'broad focus' was intended primarily as 'focus on the whole utterance' – a replacement, as we shall see shortly, for the earlier term 'normal stress'. However, it now seems clear that the important difference between broad and narrow is a matter of degree: focus can apply to constituents of any size, from individual morphemes such as prefixes to whole clauses or sentences. Sentence (4), for example, clearly expresses a contrast between the phrase *five francs* and one of the corresponding phrases in the first half of the sentence (*a sandwich*, *the car keys*, etc.), and for that reason seems like 'narrow' focus. Crucially for the discussion here, though, the broad focus problem is still present, because the contrast applies to a whole constituent rather than a single word, and only one word in the contrasted constituent can bear the main accent.

view and the highlighting view. The normal stress view goes back at least to Newman (1946). It was dominant, at least in American linguistics, until the 1970s, finding perhaps its definitive expression in the Nuclear Stress Rule of Chomsky and Halle (1968: ch. 3), and it still informs some mainstream work such as Cinque (1993) and, especially, Zubizarreta (1998). According to this view, there is one pattern of prominence ('normal stress') that can be specified by rule for every sentence. This pattern assigns a single most prominent stress – primary stress – to one word in the sentence. Normal stress has no meaning or function: it is simply the result of the operation of phonological rules on surface syntactic structures. Any deviation from normal stress, on the other hand, involves 'contrastive stress', which signals some sort of contrast or emphasis on the stressed word. Contrast or emphasis is essentially unpredictable and beyond the scope of linguistic rules – in the terms we used in chapter 1, it is paralinguistic.

The highlighting view, which seems to have its roots in the Prague School notion of 'functional sentence perspective' (e.g. Daneš 1967), was forcefully championed by Bolinger (e.g. 1958, 1972b), and was later enthusiastically taken up by Chafe (e.g. 1974, 1976) and by other broadly 'functionalist' researchers. Bolinger argued that words – any word in any utterance – can be 'focused' or 'highlighted' to signal newness, contrast, or some other special informativeness, and that focused words are marked by pitch accents. All pitch accents are individually meaningful, and no one of the pitch accents in an utterance is primary. There is thus no sharp divide between 'normal' and 'contrastive' stress: normal stress is merely one end of a continuum of informativeness. Bolinger explicitly rejected the idea that one particular pattern of pitch accents is assigned by rule, and argued that the use of pitch accents is guided by 'interest' and 'power' – factors that are, in the terms used here, paralinguistic: what speakers decide to highlight is not a matter of grammar, but a matter of what they are trying to say on a specific occasion in a specific context. His view of the attempt to provide linguistic rules for sentence stress was succinctly summed up in the title of his paper, 'Accent is predictable (if you're a mind-reader)' (1972b).

In some cases, a plausible account of broad focus can be cast in terms of the highlighting of individual words. For example, in a sentence like *He went on FOOT*, we can easily argue that *foot* is the key to the meaning of the sentence, and that, given *foot*, the verb form *went* is very largely predictable. This is true whether the context is *He left the car and went on foot* (in which case we might speak of focus on the whole constituent *went on foot*), or *He didn't go by car, he went on foot* (which clearly focuses on *foot* or on the phrase *on foot*). In either case, *foot* can be seen as the most informative word in the sentence, and hence the word most appropriately highlighted by an accent; there is no need to invoke a special concept of broad focus. But in at least some cases of broad focus, this kind of explanation appears deeply implausible. In our example *five FRANCS*, for instance, we would have to say that *francs* bears the main accent in the broad focus case because it is somehow more informative or salient than *five*. It is hard to see how this could be so in an exchange like the following:

(5) A: What did they give you for participating in the experiment?
B: Five francs.

In this case, *francs* is almost entirely predictable if the conversation takes place in a country where the unit of currency is the franc; *five* is the information of interest. Yet the main accent is on *francs*. The best the highlighting view can do in a case like this is to say that other factors are at work in determining discourse salience and therefore focus. (For example, *francs* might be more salient because it is a noun, or because it stands last in the utterance.) But the argument is essentially circular, since the main evidence for the supposed greater discourse salience of *francs* is the very fact that it is accented. Examples like (5) form the core of the case against a pure highlighting view.

6.1.2 *The Focus-to-Accent approach*

The gap between the highlighting theory and the normal stress theory narrowed somewhat during the 1970s and 1980s. This was due in large measure to the development of what we might call, following Gussenhoven (1983a), the 'Focus-to-Accent' (FTA) approach (exemplified by, among others, Schmerling 1976; Ladd 1980a; Gussenhoven 1983a; Selkirk 1984). This approach built on explorations of the notion of 'focus' in the context of generative work on syntax and semantics (beginning with Chomsky 1972 and Jackendoff 1972), and depended for its success on Pierrehumbert's rehabilitation of Bolinger's notion of pitch accent.[3] The hallmarks of the FTA view are that it distinguishes the semantic/pragmatic notion 'focus' from the phonetic/phonological notion 'accent', and – crucially – that it allows focus to apply to portions of utterances larger than an individual word. In this view, the linguistic description of

3 Similar ideas were actually developed earlier by Halliday (1967a, 1967b), but these papers unfortunately had little direct influence on mainstream American work.

sentence stress involves two complementary but essentially separate aspects: a statement about which parts of an utterance are focused, and a statement about how a given pattern of focus is conveyed by the location of the accent. The speaker's decision about which word or constituent to focus is subject to all kinds of contextual influences which are at best poorly understood: these are the factors with which Bolinger, Chafe, Halliday, and others have always been concerned. However, once we specify the focused part of the utterance (which, to repeat, can be more than an individual word), the location of accent on a specific word within the focused constituent follows more or less automatically by language-specific rules or structural principles such as Gussenhoven's 'Sentence Accent Assignment Rule' and 'Minimal Focus Rule'. Rules like these cover much of the ground that was dealt with by traditional normal stress rules.

Since the mid 1980s, something like the FTA view has been the foundation of much work on the semantics of focus (e.g. Rooth 1985, 1992 or Krifka 1991, 2006 on the interpretation of the scope of logical operators like *only* and *even*), and on how focus interacts with syntactic and phonological organisation (e.g. Selkirk 1984; von Stechow and Uhmann 1986; Steedman 1991, 2000; Büring 1997, 1999; Féry and Samek-Lodovici 2006; Erteschik-Shir 2007). One of the central questions in much of this work is how sentence stress signals broad focus, and one of the key notions is that of *focus projection*. This is the idea that, in a hierarchical syntactic structure, focus can 'project' up the tree from an accented word to a higher node that dominates an entire constituent, with the result that the entire constituent is treated as focused. Much of the syntactic and semantic literature on focus and sentence stress since the 1980s has conceived of the problem as establishing the principles that govern focus projection, including the identification of the unmarked or neutral prominence patterns that allow focus to project.

The FTA theory thus appears to represent a productive compromise between the normal stress view and the highlighting view. It accepts the premise of the highlighting view that the location of sentence stress is always in some sense meaningful, and eliminates the implication of the normal stress view that there is a fundamental difference of kind between normal stress and contrastive stress. Normal stress is reinterpreted as broad focus on the whole sentence; contrastive stress is narrow focus; and the theory explicitly allows that focus can occur on constituents of any size. At the same time, it also provides a basis for the intuition of the normal stress view that certain aspects of sentence stress are based on grammatical principles, namely the principles that determine the location of prominence when there is focus on constituents larger than a

single word. That is, the FTA view has no trouble accommodating narrow focus on individual words, but it also allows for the existence of 'unmarked' or 'neutral' patterns that specify the location of accent in cases of broad focus on whole constituents or sentences. In this way it avoids certain paradoxes of the traditional normal stress view, such as sentences that 'must have contrastive stress', like *Even a two-year-old could do that* (cf. Schmerling (1976: 49). These are a problem only for the traditional structuralist conception of a single structurally determined 'normal stress' pattern for every sentence. As soon as we accept that accent signals focus (on constituents of some size), and that the use of certain words and constructions (such as *even*) crucially involves focus on single words or other small constituents, then there is no problem reconciling data like these with the notion of 'neutral' accent.

Despite the wide acceptance of the FTA view and the idea of focus projection as a mechanism for describing broad focus, the phenomena of focus and sentence stress continue to be much debated in the literature. For one thing, there are important problems on the phonological side of the FTA view, in the sense that 'accent' is almost certainly too simple a way of describing what makes focused words and constituents prominent. These problems will occupy us extensively in chapter 7; for now I note only that these problems are the main reason I use the term *sentence stress* in this chapter. For another thing, the syntactic and semantic side of the problem is plagued both by theory-internal disagreements about the relation between information structure and truth-conditional semantics in a formal theory of grammar, and by having to work with many slippery concepts (such as 'familiarity', 'contrast', 'topic', and 'point of view') that can reasonably be discussed in a variety of incompatible ways. In all of this discussion, certain features of the Bolingerian highlighting view continue to make themselves felt.

Some of the work that attempts to dispense with the notion of focus projection has been based, roughly speaking, on drawing distinctions between different aspects of pragmatic meaning that are conflated in the term focus or in the notion of highlighting. In particular, various authors distinguish the 'newness' of a word or phrase (e.g. whether the entity referred to has been recently mentioned or is newly introduced to the discourse) from its 'contrastiveness' or 'informativeness' (e.g. whether the point of the sentence is to state that a proposition is true of one discourse entity rather than another). Halliday (1967a) drew this distinction quite clearly, and related ideas have more recently been developed by, for example, Lambrecht (1994); Vallduví and Vilkuna (1998); É. Kiss (1998); and Steedman (2000; Kruijff-Korbayová and Steedman 2003). For example, Steedman's conception of information structure involves

two orthogonal dimensions, *theme/rheme* and *kontrast*[4]*/background*: the first distinction is about which part of the sentence is being presented as new or interesting, and which as already accessible in the discourse; while the second is about how the parts of the sentence fit into the sentence's truth-conditional propositional content.

By distinguishing two aspects of the problem, most of these authors leave themselves with some way of getting at the problem of broad focus. For example, Steedman's idea that the proposition is articulated into theme and rheme (and specifically the idea that theme and rheme can be as small as a single word or as large as a whole sentence) gives his theory a means of dealing with many of the phenomena that have been discussed under the heading of focus projection. On the other hand, the semantic/pragmatic notion of kontrast shares a great deal with Bolinger's notion of focus as highlighting, and, like Bolinger, Steedman (again following Vallduví and Vilkuna) treats kontrast as being signalled by accent. It is therefore not clear that he can entirely avoid the problem of having to assign accent only to single words rather than to constituents. Similar comments seem to apply to other work that explicitly does away with the idea of focus projection.

Another reason for the persistence of Bolingerian ideas is that there are important areas of empirical investigation where the issue of broad focus does not really arise. One such area deals with the dynamic organisation of conversation and the flow of information between speaker and hearer. There are a number of largely independent traditions that study a number of specific issues, including: the different kinds of discourse statuses that entities may have (previously mentioned, newly introduced, etc.; e.g. Prince 1981); how long such discourse statuses last and what causes them to change (e.g. in order to count as previously mentioned, must the earlier mention have taken place within a certain amount of elapsed time? a certain number of turns? a certain number of topic shifts?; e.g. Grosz and Sidner 1986); the way in which entities are referred to depending on their discourse status (a full descriptive noun phrase, a noun with a definite article, a pronoun, etc.; e.g. Halliday and Hasan 1976; Gundel, Hedberg, and Zacharski, 1993; Couper-Kuhlen and Selting 1996); and, finally, the effect of discourse status on whether referring expressions are accented or not and how they are arranged syntactically (e.g. Nooteboom and Terken 1982; Nooteboom and Kruyt 1987; Terken and Hirschberg 1994; Clifton and

4 *Kontrast* – sometimes read aloud as 'K-contrast' – is spelled that way in order to identify it as a technical term with a formally definable meaning, distinct from the ordinary-language word *contrast*; cf. Vallduví and Vilkuna (1998).

Frazier 2004; Swerts, Krahmer, and Avesani, 2002). What is common to all these traditions is a concern with the reasons for which a speaker decides to focus on a constituent or not. Because they tend to study single entities that can be referred to with single accented or unaccented words, their attention is concentrated on the relation between the discourse status of words and the accentuation of words. The problem of where to locate accent *within* a focused constituent does not arise, and broad focus therefore receives little attention.[5]

Finally, as with the description of pitch patterns, so in the description of prominence patterns the Bolingerian emphasis on intonational universals is never very far from the surface. One area in which the FTA and highlighting views of broad focus clearly make different predictions is in cross-language comparison. If accents are directly meaningful signals of focus or discourse salience, unpredictable except with reference to speakers' intentions and specific contexts, and if broad focus is to be explained in the same terms as narrow focus, it is easy to argue that the relation between accent and focus is part of some universal (and possibly prelinguistic) intonational highlighting function. This was very explicitly claimed by Bolinger, and is implicit in the views of many who study accent in discourse. By contrast, those who argue for the existence of focus projection, or otherwise admit a role for structure in discussing broad focus, often invoke language-specific rules or principles. In the remainder of this chapter, therefore, I survey sentence stress data from a dozen or more languages, with the aim of showing that a universal highlighting theory with a direct relation between accent and focus is empirically inadequate.

5 In this context it is also important to mention that the term *focus* is used in two essentially incompatible ways in the recent literature. In the tradition that begins with Grosz and Sidner (1986), a discourse entity is said to be 'in focus' if it is the current topic of conversation, that is, if it is the most salient or activated in the speakers' awareness. Such entities are obviously 'given' rather than 'new', and as such are likely to be referred to with *unaccented* expressions in English. This usage of the term contrasts with the older usage originated by Bolinger, which is carried forward into both the formal semantics tradition of, for example, Jackendoff (1972) and Rooth (1985), and the intonational work of, for example, Ladd (1980a), Gussenhoven (1983a), and Selkirk (1984). In this usage, focus attaches to the most informative parts of the sentence, which are accordingly likely to be pronounced *accented* in English. Though these two usages seem like exact opposites, the confusion persists, because the older usage is itself multifaceted. In particular, work on focus in the older sense does not always distinguish the logical and truth-conditional aspects of focus from the discourse-level pragmatic aspects such as newness, givenness, and informativeness, or predictability more generally. The latter aspects are obviously related to the Grosz–Sidner sense of focus, which feeds the confusion and the cross-fertilisation. Readers of the literature on focus need to be alert to the possibilities of misunderstanding due to the multiple meanings that have been attached to the word.

Specifically, I aim to demonstrate the existence of language-specific differences in broad focus patterns of sentence stress by considering comparable sentences in different languages. Consider, for example, hypothetical data of the following sort:

	Language A	*Language B*
	This is book RED	This is BOOK red
	I bought car NEW	I bought CAR new
	He has nose BIG	He has NOSE big

In language A alone, the highlighting theory could argue that, in the absence of other influences, the main accent goes to the end of the sentence (Bolinger's 'accents of power', widely observed in many contexts). In language B alone, the theory might suggest that nouns are intrinsically more informative than adjectives and are thus likely to attract the speaker's attention (Bolinger's 'accents of interest'). But the *comparison* between languages A and B makes it impossible to maintain that these principles are universal, or that conflicts between these principles are resolved by individual speakers in individual situations on the basis of universal principles. The difference between the two languages seems to point inevitably to language-specific – and hence possibly structure-dependent – principles for placing accent within broad focus constituents. This comparative approach – which to my knowledge was first suggested in a 1986 paper by Michel Kefer that was unfortunately never published – is the general strategy adopted in the following section. On the basis of such comparisons, it is possible to demonstrate – conclusively, in my view – that there are consistent cross-language differences in patterns of sentence stress. This in turn means that, whatever exactly the relation between focus and accent may be, it is not simply a matter of applying some universal highlighting gesture to individually informative words.

It is important to make clear that in comparing isolated citation forms of sentences we are not simply reverting to the old-fashioned notion of normal stress. As we saw in section 6.1.1, normal stress was seen by structuralist and early generative writers as a single structurally determined prominence pattern to which context is by definition irrelevant. In context, one might get 'contrastive stress' anywhere in a sentence, but in principle one ought to be able to elicit the normal stress pattern without any context at all – that is, as citation forms of sentences. If we reject this notion of normal stress in favour of the FTA view, we have to be suspicious of eliciting prominence patterns out of context. It has often been suggested (e.g. by Bolinger 1972b) that it is methodologically unsound to try to determine how sentences are accented 'out of context' – for instance, when read as examples in a linguist's study – because speakers may imagine all kinds of contexts which will affect sentence stress in unpredictable ways.

The methodological aspect of this general line of argument obviously has some force: speakers may indeed imagine contexts which will affect sentence stress, and we must be careful not to base theories of sentence stress on data that may be distorted in this way. On the other hand, it is worth emphasising that with the FTA view we do not define anything in terms of contextlessness. The notion of normal stress is redefined as the prominence pattern that can convey broad focus: there is no claim that this pattern is contextless, but only 'unmarked' – the pattern that is chosen when there is no compelling grammatical or contextual reason to choose some other. We continue to make reference to the communicative intention of the speaker and the context in which the utterance is used: the broad focus pattern *five FRANCS* is appropriate in a wide range of contexts, while the 'narrow focus' pattern *FIVE francs* is appropriate in very few contexts, but nothing depends on any assumption of 'contextlessness'. This represents an important change of orientation from the structuralist and early generative view.

Obviously we need to be wary of citation-form sentence stress patterns: because of the methodological problems just cited, we may not succeed in eliciting broad focus patterns as citation forms. Yet in a sense this matters less for cross-language comparison than for the investigation of a single language. If we compare citation forms across languages and find consistent differences in the patterns of sentence stress, it is difficult to explain away these differences with reference to theoretical or methodological problems with elicited citation forms. That is, it is difficult to imagine universal pragmatic principles that would account for consistent differences between languages in structurally parallel citation forms. I believe these comparative data show that we must go beyond the highlighting theory and adopt some version of the FTA view.

6.2 Sentence stress patterns across languages

The discussion of cross-language differences is organised as follows. First, I present three general types of cases in which I believe the evidence for cross-language differences is fairly clear. These are:

(1) questions;
(2) deaccenting and 'semantic weight';
(3) predicates versus arguments.

Following an extensive presentation of the data, I briefly discuss the ways in which these three general types of cases might be interrelated, and I consider which aspects of sentence stress, if any, are governed by universal principles.

6.2.1 *Questions*

6.2.1.1 Yes–no questions

The most convincing cross-linguistic differences are those seen in yes–no questions (YNQs). In some languages, like English, YNQs are treated exactly like statements for purposes of sentence stress rules. Thus in 'citation form' we would expect

(7) Statement: She bought a BOOK.
YNQ: Did she buy a BOOK?

Both of the following forms are distinctly non-neutral in some way:

(8) Statement: She BOUGHT a book.
YNQ: Did she BUY a book?

English YNQs, like English statements, *can* have the main accent on the verb, but normally only if the verb is not followed by a lexical noun:

(9) Statement: She's SLEEPING.
YNQ: Is she SLEEPING?

In short, there is nothing in these data to suggest that the distinction between statements and questions is relevant to sentence stress.

In other languages, however, such as Russian, the two sentence types pattern differently. Russian statements, like English statements, have the greatest prominence on the noun if there is one following the verb:

(10) Ona kupila KNIGU (lit. she bought BOOK)

In YNQs, on the other hand, the neutral prominence pattern or citation form has the greatest prominence on the verb, regardless of whether the verb is followed by a lexical noun:

(11) (a) Ona SPIT? (lit. she SLEEPS?)
(b) Ona KUPILA knigu? (lit. she BOUGHT book?)

A YNQ with the main accent on a noun is distinctly non-neutral; that is, the question

(12) Ona kupila KNIGU?

is felt to focus narrowly on the book in the way that English *Did she BUY a book?* focuses on the buying. If we want to say something in Russian comparable to English *Did she BUY a book?* we must change the word order to (13):

(13) Ona knigu KUPILA?

The method of comparing similar sentences across languages effectively precludes arguments about the nature of neutral sentence stress or the exact contexts in which one prominence pattern or another would be appropriate. The differences must be seen as genuine.[6] Languages like Russian accent YNQs on the verb, regardless of whether the sentence also contains nouns; languages like English accent the verb in YNQs only if there is no following lexical noun. In this case, 'languages like English' include all the Germanic and Romance languages, with the exception of Romanian and the probable exception of Southern Italian; 'languages like Russian' include Romanian, Hungarian, Greek and many, if not all, of the Slavic languages. For most languages, the facts remain to be determined.

The existence of two neutral or default sentence stress patterns for YNQs can be related to cross-linguistic facts about question particles. In languages that have question particles, there are two main locations for the particle. In one type of language, the question particle occurs at the edge of the sentence, either the beginning (e.g. Yoruba) or the end (e.g. Chinese); in these languages there is no relation between the location of the particle and the location of focus. In the other type of language, the particle attaches to the focus of the question, and if there is no special focused word, then the particle – like the main accent in Russian – attaches to the finite verb. In Turkish,[7] for example, the question particle normally occurs as one of the suffixes or enclitics on the finite verb or other predicate:

(14) (a) Gazete geldi mi?
'Did the newspaper come?' (lit. newspaper come-PAST INTERROG.)
(b) Yorgun musunuz?
'Are you tired?' (lit. tired INTERROG.-2PERS.-PLURAL)

6 In his last letter to me (10 June 1990), Bolinger conceded that this case 'is a puzzler, certainly, for what I would want to claim', though I think it is fair to say that he nevertheless remained convinced of the ultimate validity of the highlighting view that he had held for so long.

7 The accentual data from Turkish here and below are based on class notes from a structure-of-Turkish seminar held at Cornell in 1976–7. Most of the sentences are from Underhill 1976, and the patterns of sentence stress shown are based on the speech of Vedia Ceranoğlu, the seminar's native-speaker consultant.

However, if there is a focus on some specific word or constituent other than the verb, the question particle is attached to the focus:

(15) (a) Mehmet mi geldi?
'Was it Mehmet who came?'

(b) Buraya uçakla mı geldiniz, vapurla mı?
'Did you come here by plane or by steamship?'
(lit. here-to plane-by INTERROG. come-PAST-2PERS.-PLURAL, steamship-by INTERROG.)

In a number of other languages, the question particle immediately follows the first word or constituent of the sentence, and a specially focused word may move to first position; if there is no specially focused word, the word in first position is once again the finite verb. The following examples are from Russian:

(16) (a) Mark li čitaet?
'Is MARK reading?' (lit. Mark INTERROG. reads)

(b) Čitaet li Mark?
'Is Mark reading?'

Analogous examples are possible in other languages, including Classical Latin. Once again, there seems to be a special connection between the focus of a question and the finite verb.

6.2.1.2 WH-questions

For YNQs, then, it appears that there are two different kinds of sentence stress rules cross-linguistically: one in which questions exhibit a special neutral location for the main accent (on the verb), and one in which the neutral location of the main accent is similar in questions and statements. Similar conclusions seem to hold for questions containing question words like *who*, *how*, *what* (usually called WH-questions in English, and here abbreviated WHQs). Various works on focus and accent (in particular, Culicover and Rochemont 1983: 139–44) deal uneasily with the prominence of the WH-words in WHQs. Logic seems to suggest that the WH-word is the focus of the question, and yet, in English at least, the WH-word does not normally bear the most prominent accent. That is, English has

(17) Where are you GOING?

rather than

(18) WHERE are you going?

though the latter would seem to be demanded by a purely focus-based account of accent placement.

Languages do exist, however, in which the WH-word is the most prominent in the unmarked sentence stress pattern. This appears to be the case, first of all, in languages without WH-movement, where the WH-word is normally the last noun phrase of the sentence. Thus in Turkish we find:

(19) Halil'e NE verdiniz?
'What did you give to Halil?' (lit. Halil-to WHAT you-gave)

And in Bengali we have:

(20) Ram KAKE dekhlo?
'Whom did Ram see?' (lit. Ram WHOM saw)

Similarly, English WHQs without WH-movement – most often occurring as echo questions – have the main accent on the WH-word:

(21) (a) You did WHAT?
(b) They went WHERE?

and so on.

More relevantly for the comparison with ordinary English WHQs, there are languages with WH-movement that do put the nuclear accent on the WH-word, so long as the sentence is fairly short. In Romanian, for example, one says:

(22) (a) UNDE mergi? 'Where are you going?'
(b) CÂŢI bani ai? 'How much money do you have?'
(c) CÂND a plecat? 'When did it leave?'
(d) CINE a chemat? 'Who called?'

Essentially the same is true of Hungarian:

(23) (a) KI az? 'Who is that?'
(b) MIT vettél? 'What did you buy?'
(c) MILYEN volt a vacsora? 'How was the dinner?'

and of Greek:

(24) (a) PU ine? 'Where is it?'
(b) JIATI efije? 'Why did she leave?'
(c) TI idhes? 'What did you see?'

As with YNQs, then, we have two basic typological patterns, one in which WHQs follow the same sentence stress principles as other sentence types (as in English), and one in which there is a special rule for WHQs, whereby the neutral location for the main accent is on the WH-word (as in Romanian or Greek).

6.2.1.3 A note of caution

Before leaving the topic of sentence stress in questions, it is important to draw attention to two potential difficulties in determining what the facts are. The first of these is that there are effects of sentence length that affect simple generalisations of the sort just stated; the second is that in many cases, especially in questions, it is not a straightforward phonetic task to identify the most prominent word or syllable. I have tried to avoid these problems wherever possible in the discussion in this section, but it is important to sketch the issues before going on.

Briefly considering the first problem, we note that in WHQs in languages that are said to put the main accent on the WH-word, there may be one or more accents later in the sentence if the sentence is long. Thus in Romanian one would be likely to find either version of (25) and (26):[8]

(25) (a) Unde ai cumpărat cravata ASTA?
(b) UNDE ai cumpărat cravata asta?
'Where did you buy that necktie?'

(26) (a) Cu cine ai vorbit la FACULTATE?
(b) Cu CINE ai vorbit la facultate?
'Who did you talk to at the university?'

The issue of sentence length and its effect on sentence stress is discussed further in section 7.3.1.

Turning now to the more substantive question of how to identify the most prominent syllable, let us consider the claim that YNQs in many languages have the main accent on the verb. The Russian data on this point are fairly clear, because the normal YNQ intonation in Russian involves a high accent peak on the most prominent syllable: L+H*_L, or something of the sort. (Normal statement intonation is !H*_L or H+L*_L; i.e. the question peak is aligned later than the statement peak; cf. Odé 1989: ch. 5; Makarova 2007.) In many other Eastern European languages, however, the normal YNQ intonation is L*_H_L% (see sections 2.5 and 4.1.4). Consequently, in these languages the syllable that is described here as 'most prominent' is phonetically *low*, and is followed later in the sentence by a peak-and-fall which is often acoustically more salient, at least to Western European ears, than the L*. This analysis of the low syllable as nuclear or most prominent, and the peak-and-fall pitch movement as post-nuclear and subordinate, ought to be uncontroversial – as

8 In fact, as can be heard from the sound files for these examples, it is rather difficult to distinguish one of these versions from the other.

noted in section 2.5, it has been stated clearly by Hungarian linguists at least since the 1960s. However, some researchers who are concerned fairly narrowly with phonetic realisation (e.g. Gósy and Terken 1994; Xu 2005) still equate pragmatic prominence with acoustic salience, and reject or ignore the idea that a low-pitched syllable might be abstractly and pragmatically more prominent than the following high peak. This being the case, it is important to point out that the typological statements just made in this section depend on the validity of the phonological description of the contours under discussion. Extensive phonetic evidence for the phonological description, at least in Greek, is presented by Arvaniti, Ladd, and Mennen, (2006), who criticise Xu's revival (Xu 2005) of a strongly universalist approach to focus and question intonation.

A similar problem arises in the case of the early main accent on WHQs. Phonetically, there is considerable similarity between the contours on, for example, the Romanian and Italian phrases meaning 'Where are you going?':

(27) (a) H* L L%
Romanian: U n d e m e r g i ?
(b) H* H+L* L L%
Italian: D o v e v a i ?

However, the discussion in section 6.2.1.2 above is based on an analysis in which the phonological structure of the two is different: in the Romanian case the nuclear accent is on the WH-word *unde*, and *mergi* is post-nuclear; in the Italian case *dove* is prenuclear, and there is an H+L* nuclear accent on *vai*. (That is, the Italian tune contains the same accent type used in broad focus statement intonation in the Romance languages generally; see section 3.1.3 and figure 3.7.) Again, the typological statements are valid only if the phonological analysis is correct. This case is not as well-studied as the YNQ contour of Greek and Hungarian, but considerations of both phonetic detail and native-speaker intuition suggest that the analysis assumed here is correct.

The phonetic considerations involve the course of F_0 on the unstressed second syllable of the WH-word (e.g. *unde* or *dove*). In Romanian (27a) the pitch comes down quickly on the unstressed syllable of *unde*, and there is no further movement on *mergi*. This suggests that there is only one accent on the phrase, on *unde*. In Italian (27b), by comparison, the pitch on *dove* stays relatively high and steps down to the following stressed syllable *vai*. The distinction between the two contours can be seen more clearly if we increase the number of unstressed syllables between the stressed syllable of the WH-word and the stressed syllable of the verb, as, for example, in the following sentences

meaning 'Where did you buy it?' (the lexically stressed vowel of the verb is underlined):

(28) (a) H* L L%
Romanian: Unde l-ai cumpărat?
(b) H* H+L* LL%
Italian: Dove l'hai comprato?

As for the matter of native-speaker intuitions, informal questioning of speakers of various languages reveals a clear difference: Romanian or Hungarian or Greek speakers have firm intuitions that the WH-word bears the main accent in sentences like these, while Italian or Portuguese speakers tend to be uncertain. Obviously, the situation would be clearer if Italian or Portuguese speakers had clear intuitions that the verb bears the main accent, but the difference is still noteworthy. In this connection it is worth considering that ordinary speakers of European languages generally seem inclined to equate high pitch with sentence stress, if asked to make metalinguistic judgements about which word is most prominent (see e.g. Bolinger 1958: 131–4, reprinted 1965: 38–43). This may explain the uncertainty on the part of native speakers of Italian or Portuguese about the location of 'the' main accent in WHQs: the WH-word has the highest pitch, even though, on the analysis proposed here, it is not the most prominent accent of the sentence. In any case, with WHQs, as with YNQs, the typological discussion of sentence stress depends to some extent on the validity of the phonological analysis.

Finally, note that Kefer (1986) attempted to sidestep the problem of justifying the phonological analysis by using a pre-theoretical phonetic definition of sentence stress and by concentrating his typological study on statements. While such an approach may be more convincing in the short run, I believe it is ultimately necessary to acknowledge that our conclusions depend on our phonological analysis, which is why I have discussed the matter at some length here. In the long run, cross-linguistic experiments exploring native-speaker intuitions may help to establish the reality of the different patterns of relative prominence described in the foregoing paragraphs.

6.2.2 *Deaccenting and 'semantic weight'*

The data on sentence stress in questions, as I have just presented them, are clearly difficult to reconcile with a universalist highlighting theory. Cross-linguistically there seem to be two distinct types of prominence patterns for both YNQs and WHQs, and there are clear parallels between the possible locations of the main accent in YNQs and the possible locations of question particles.

All these facts point to the conclusion that accentuation in questions is a matter of the grammar of specific languages rather than of universal principles of highlighting focused words. On the other hand, the data are admittedly limited, and potentially questionable in some cases. Sentence stress has not been studied in questions anywhere near as much as in statements, and it is entirely possible that there are complexities beyond the brief presentation here that could be used to support a different interpretation. This section, therefore, concentrates on well-studied sentence stress data from declarative sentences that form the heart of the case for the highlighting view.

Specifically, in this section we consider cases in which, in English, sentence stress can be influenced by the relative informativeness of words or constituents in a sentence. For example, it is well known that the main accent tends *not* to be placed on elements that are repeated or 'given' in the discourse, or on elements that are vague or generic. For adherents of the highlighting view, this fact is a clear illustration of the general principles governing sentence stress in any context: the speaker assesses the relative semantic weight or informativeness of potentially accentable words, and puts accent on the most informative point or points of the sentence. However, it is not hard to show that these observations, valid though they may be for English, do not apply in all languages. To back up the claim of non-universality, this section adopts Kefer's approach of considering structurally parallel examples from different languages, as we did in the discussion of questions. The discussion is divided into three sections, each dealing with one of three subcases: contextual deaccenting; indefinite pronouns; and 'semantically empty' words. In every instance the existence of cross-linguistic differences seems beyond dispute.

6.2.2.1 Contextual deaccenting

An important accentual difference between languages involves the treatment of repeated words or phrases, and more generally the treatment of 'given' information. A few examples from English follow:

(29) (a) A: I found an article for you in a German journal.
B: I don't READ German.
(b) I brought her a bottle of whisky, but it turns out she doesn't LIKE whisky.

In both of these, a word that we might expect to bear the main accent (*German*, *whisky*) fails to do so in a context where it has recently been used, or where the entity to which it refers has recently been mentioned. In my own earlier work on this topic (Ladd 1980a), I used the term 'deaccenting' to describe this phenomenon, and this term has been widely adopted.

The topic of deaccenting looms large in studies of accent and focus based on the West Germanic languages, particularly English. It represents crucial evidence for those who believe that accents occur on new or otherwise salient parts of the utterance more or less without regard to structure – the highlighting view. However, there are many languages in which utterances like those in (28) do not have modified sentence stress patterns to reflect the repetition or 'givenness' of the normally accented word.

Within English, the absence of deaccenting has been noted in both Hawaiian Pidgin (Vanderslice and Pierson 1967) and in Indian and Caribbean English (Gumperz 1982: ch. 5). Vanderslice and Pierson give the following example (their example 8, with spelling standardised):

(30) Forty-three per cent is government OWNED and fifty-seven per cent is privately OWNED.

One of Gumperz's examples from Indian English (example 21, p. 125, with considerable discussion of prosodic detail omitted) is:

(31) If you don't give me that CIGARETTE I will have to buy a CIGARETTE.

Looking beyond English, Cruttenden (1993) reported preliminary results from a study of this and other cross-language differences of sentence stress patterns. His findings were that some languages (like English) more or less insist on deaccenting repeated material, while others (like Spanish, in his sample) quite strongly resist it. This difference has since been confirmed experimentally (specifically, for Dutch and Italian) by Swerts, Krahmer, and Avesani 2002.

I have elsewhere (Ladd 1990a) given examples from both Romanian and Italian, which are certainly to be classed among the languages that resist deaccenting. One such case was the following Romanian sentence, which was spoken by a university employee who had come to inspect the contents of the apartment I was vacating after my year as a Fulbright scholar, and to check them against an inventory list that I had signed at the beginning of the year:

(32) [. . . o să vedem] ce AVEŢI şi ce nu AVEŢI
(lit. [. . . we'll see] what you-have and what not you-have)

An idiomatic translation that faithfully preserves the accent pattern would be something like (33a) instead of the expected (33b):

(33) (a) (?) So let's see what you HAVE and what you don't HAVE.
(b) So let's see what you HAVE and what you DON'T have.

The following example from Italian is particularly convincing for speakers of Standard English. The speaker was former Italian President Scalfaro, on the subject of the judicial investigations into massive bribery and corruption during the so-called 'Tangentopoli' scandal of the early 1990s:

(34) [le inchieste] servono a mettere a POSTO cose andate fuori POSTO
(lit. [the investigations] serve to put to place things gone out-of place)

Again, an idiomatic translation preserving the prominence pattern (35a) sounds decidedly odd; with such a rhetorical parallelism, the pattern in (35b) is more natural in English:

(35) (a) The investigations are helping to put back in ORDER things that have got out of ORDER.
(b) The investigations are helping to put BACK in order things that have got OUT of order.

Even in languages that resist deaccenting, the main accent *can* be shifted away from a neutral or default location under certain circumstances. In Italian or Romanian, which in general are strongly non-deaccenting, explicit metalinguistic corrections can have paired accents on the corrigendum and the correction, irrespective of word order:

(36) Non ho detto CASA bianca, ho detto COSA bianca
'I didn't say white HOUSE, I said white THING'

Italian also fairly readily allows contextual deaccenting of non-finite verb forms, and of predicate noun and adjective phrases, especially where the resulting main accent is on an auxiliary, and especially in negative sentences:

(37) (a) Non È la mia bici (l'ho presa in prestito).
'It's NOT my bike (I borrowed it).'
(b) Non È intelligente.
'He's NOT intelligent.'
(c) Non ti POSSO aiutare.
'I CAN'T help you.'
(d) Non c'HA invitato.
'He DIDN'T invite us.'

But there are clear syntactic restrictions on such accent shifts. I once encountered the following sentence in reading a children's story aloud in Italian (in the story, a young zebra is being instructed how to run):

(38) Correre è come camminare in fretta, soltanto si deve andare molto più in fretta.
'Running is like walking fast (lit. in haste), only you have to go much faster (lit. much more in haste).'

I used the sentence stress pattern in (39a), based on the pattern that would be appropriate on the literal translation, shown in (39b):

(39) (a) Correre è come camminare in FRETTA, soltanto si deve andare molto PIÙ in fretta.
(b) Running is like walking in HASTE, only you have to go much MORE in haste.

This pattern is unhesitatingly rejected by Italian native speakers, apparently because it deaccents only a part of the adverbial phrase *molto più in fretta*.

Another type of deaccenting that is not uncommon in English, but which never occurs in the Romance languages, involves the actual modification of lexical stress. A classic example is Bolinger's *This whisky wasn't EXported, it was DEported* (1961b: 83). An attested example from a BBC news broadcast is the following:

(40) Greek divers have found the wreck of the British liner BriTANNic, sister ship of the TItanic ...

in which the lexical stress on *Titanic* is shifted to focus on the point of contrast between the two ships' names. A similar case involves the English *-teen* numbers, which have their lexical stress on *-teen* in citation form, but can readily be observed with the stress shifted in counting:

(41) FIFteen, SIXteen, SEVenteen, EIGHTeen, NINEteen, TWENty

Corresponding shifts of lexical stress in the Romance languages are impossible. In speaking Italian I once attempted the following utterance:

(42) Moglie quaranTENne, marito CINquantenne
'Forty year old wife, fifty year old husband'

shifting the normal lexical stress on *cinquanTENne* ('fifty year old') to the first syllable in order to emphasise the contrast between 'forty' and 'fifty'. The result, like my attempted deaccenting in (39), is completely unacceptable in Italian. I have elsewhere (Ladd 1990a) noted the Romanian analogue of the English teen-counting series, namely counting by tens to 100:

(43) cinZECI, şaiZECI, şapteZECI, optZECI, nouăZECI, o SUtă
'50, 60, 70, 80, 90, 100'

The counting context is the same: the morphological transparency of the English teens and the Romanian multiples of ten is strikingly similar; but in Romanian there is no tendency whatever to shift the stress away from the repeated element *-zeci*. (Note that orthographic *-zeci* represents a single syllable [zɛtʃ].)

While the Romance languages mostly do not allow direct deaccenting, they all have a number of morphosyntactic strategies for achieving similar intonational effects. The most common of these is right-dislocation, in which a constituent (usually, but not always, a noun phrase) is moved outside the clause, leaving some sort of pronoun in its place. In these cases, the right-dislocated constituent is always pronounced as an intonational tag (see section 7.2.2), which implies low pitch in ordinary statement intonation, and the phonetic effect of right-dislocation is thus very similar to the phonetic effect of deaccenting the last word in an English sentence. The following example from Italian illustrates both the syntactic device and the similarity of the phonetic effect to English deaccenting. The sentence was spoken to a child whose baby brother had just had his evening bath:

(44) Adesso faccio scorrere il TUO, di bagnetto.
(lit. now I-make run the yours, of bath-DIM.)

In English, in a comparable context, we would certainly have:

(45) (a) Now I'll run YOUR bath.

A literal rendition of this in Italian would be

(45) (b) Adesso faccio scorrere il TUO bagnetto

which is possible, but considerably less natural than (44). Right-dislocation used in this way is also an extremely common strategy for 'achieving deaccenting' in French (see section 3.3.2).

This brings up the question of the relation between accentuation and word order. Vallduví (1991; see also Vallduví and Zacharski 1994), basing himself primarily on comparisons between Catalan and English, proposes that some languages (like English) have 'plastic' prominence patterns, and that other languages (like Catalan) do not. A language with a non-plastic pattern is constrained to vary word order to shift words into sentence locations where they will appear with or without accent as appropriate. This word-order variation can be accomplished either directly by 'scrambling', or by the use of marked syntactic structures such as right-dislocation, clefting, fronting, and so on. Not all Romance languages are equally non-plastic: Catalan seems to be the most

strongly resistant to moving the main accent out of phrase-final or sentence-final position, and detailed comparison clearly suggests that languages resist deaccenting to different degrees. For example, Enric Vallduví (personal communication) states that the Italian sentences involving accent on auxiliaries in (37) have no direct counterparts in Catalan, which must use right-dislocation in all such cases.

In any case, the conclusion from this brief survey must be that there is a difference between deaccenting and non-deaccenting languages – or more precisely, between (a) languages that permit, prefer, or virtually require the deaccenting of repeated and otherwise given material, and (b) languages in which such deaccenting is dispreferred or syntactically restricted, allowable strictly in cases of metalinguistic correction, and/or achievable primarily through word-order modifications. This conclusion is difficult to reconcile with the claim that accent is used universally to highlight focused words.

6.2.2.2 Indefinite pronouns

Another specific point of disagreement between FTA and highlighting accounts of accentuation involves 'indefinite pronouns' such as *someone* and *nothing*. In English or Dutch, these often occur unaccented in positions where ordinary noun phrases would be accented:

(46) (a) English
  (i) They've DISCOVERED something.
    (cf. They've discovered the DRUGS.)
  (ii) She can't EAT anything.
    (cf. She can't eat FISH.)
 (b) Dutch
  (i) Ze hebben iets GEVONDEN. (lit. they have something found)
    (cf. Ze hebben de DRUGS gevonden.)
  (ii) Ze kan niks ETEN. (lit. she can nothing eat)
    (cf. Ze kan geen VIS eten.)

To the extent that such pronouns form an identifiable lexical class, it is tempting for a structure-based account to make special provision for them in sentence stress rules on the basis of their 'part of speech'. Equally, however, the semantic vagueness or indefiniteness of these pronouns seems consistent with the highlighting approach: they are unaccented because they contribute little semantic weight or interest. On the basis of data from English or Dutch alone, it is impossible to resolve this disagreement.

Cross-language comparison, though, suggests consistent differences in the way these items are handled. In some languages, indefinite pronouns are treated

for accentual purposes like any other noun phrase. In other languages, they are less accentable than ordinary lexical noun phrases. Some languages make a distinction between negative and non-negative indefinites, with only non-negative indefinites being given special treatment (Kefer 1986). For more general discussion of the accentual properties of indefinite pronouns, see Haspelmath 1997, especially section 5.7.

English treats negative indefinites rather like lexical noun phrases, while it treats non-negative indefinites as similar to personal pronouns. Thus we have:

(47) (a) I saw NOBODY
    (cf. I saw MARIA)
 (b) She HEARD something
    (cf. She HEARD it; She heard FOOTSTEPS)

In Italian, on the other hand, all indefinites are like other noun phrases, and no distinction is made between negatives and non-negatives. Thus we have:

(48) (a) Ho sentito MARIA.
    'I heard Maria.'
 (b) Ho sentito QUALCUNO.
    'I heard someone.'
 (c) Non ho sentito NESSUNO.
    'I heard nobody.'

The only interpretation for a sentence such as *Ho SENTITO qualcuno* is as a metalinguistic correction or contrast (i.e. 'I didn't see someone, I heard someone'). Nevertheless, in certain syntactic contexts, we may also detect a difference between negative and non-negative indefinites in Italian, a difference that indirectly affects the sentence stress. Specifically, the difference may be relevant to acceptability judgements about word order in Italian. Either of the following orders is acceptable:

(49) (a) Ho sentito qualcuno parlare.
    'I heard someone talking.'
 (b) Ho sentito parlare qualcuno.
    (lit. I heard talk someone)

but the verb-final order is much less acceptable with the negative indefinite:[9]

9 There is no sound file for (50a) because the speaker found the sentence impossible to say without putting the main accent on *nessuno*, and found that even with that accent pattern the sentence was extremely unnatural.

(50) (a) (?) Non ho sentito nessuno parlare.
'I heard no one talking'.
(b) Non ho sentito parlare nessuno.
(lit. I heard talk no one)

These are not, of course, just facts about word order, but have the consequence that *qualcuno* 'someone' may or may not occur in the accent-bearing location at the end of the sentence, whereas *nessuno* 'no one' must occur there. In some sense, then, negative indefinites are more likely to bear the main accent than non-negative indefinites – in Italian as in English. But the grammatical mechanism by which this comes about is different in the two languages: the notion of 'accent-bearing location at the end of the sentence' implicitly presupposes Vallduví's idea that some languages can modify their sentence stress pattern ('plastic'), while others cannot ('non-plastic') and must modify their word order instead.

6.2.2.3 'Semantically empty' content words

The third type of case in which relative informativeness seems to affect sentence stress involves intrinsically vague or general content words, such as *person*, *man*, *thing*, *stuff*, and so on. In English these often occur unaccented, again in ways that seem to favour the highlighting view. Bolinger (1972b: 636) suggests that 'the [semantic] emptiness of certain nouns can be illustrated by comparing them with other nouns that are semantically richer'. His examples include pairs like:

(51)(a) (i) He was arrested because he KILLED a man.
(ii) He was arrested because he killed a POLICEMAN.
(b) (i) I've got to go SEE a guy.
(ii) I've got to go see a FRIEND.
(c) (i) I'm going over to the DOCTOR'S place.
(ii) I'm going over to the doctor's BARN.

He rightly criticises Bresnan (1971: 271) for suggesting that nouns like *man* and *guy* form a category of 'semi-pronouns'. As Bolinger (1972b: 636) points out, this explanation is circular, because 'the only way ... to identify such nouns is by their behavior under accent'. Moreover, he notes that these forms often exhibit variability: 'where the accentual behavior with true pronouns is predictable, that of empty nouns is only highly probable'. All of this evidence seems to weigh heavily in favour of the highlighting view.

A related type of case where semantic emptiness and predictability seem to influence accentuation involves Adjective+Noun phrases. Bolinger (1972b: 638) cites several examples in which the adjective rather than the noun is accented:

(52) (a) I like it because it has a SILKY sheen.
(b) ... I don't see how you could make it to OUR place in 45 minutes unless you went through every RED light.[10]

In the same vein, Monaghan (1991: 149f.; 1992: 155ff.) identifies several combinations of Adjective+Noun that are likely to be accented on the adjective, including phrases with fairly unspecific nouns such as *meeting* and *committee*, and those with deictic or ordinal adjectives such as *latter*, *second*, and *alternative*. A few examples of such Adjective+Noun combinations taken from passages elsewhere in this book are:

(53) (a) [the motivation is] to avoid the seemingly unverifiable speculation of IMPRESSIONISTIC work, and to permit the use of familiar parametric STATISTICAL approaches (p. 19)
(b) a SIMILAR case [involves the English *-teen* numbers] (p. 234).

A final type of case in which sentence is often said to depend on factors of informativeness involves sentence-final adverbs and prepositional phrases. For example, phrases denoting places and times that are effectively here and now often occur unaccented:

(54) (a) I saw an ACCIDENT today.
(b) There's a FLY in my soup.

Chafe (1974, 1976) relies extensively on such cases in developing an account of how sentence stress, pronominalisation, and various other phenomena depend on the speaker's assessment of what is likely to be in the hearer's consciousness or at the centre of the hearer's attention (cf. Grosz and Sidner 1986; Gundel, Hedberg, and Zacharski 1993). Bolinger points out, though, that it is not just a matter of what is in the addressee's consciousness, but of more general aspects of informativeness and predictability. As he notes, 'in the sentences

They STRANGLED him to death.
They hounded him to DEATH.
They scared him to death.

10 Many readers have found this example strange, and the accent pattern on *red light* may not be a very good illustration of Bolinger's point, except diachronically: I suspect the speaker of this (attested) utterance spoke an American variety in which the phrase is lexicalised with compound stress, i.e. with main stress on *red*.

(once more answering *What happened*?), the first de-accents *death* because strangulation normally involves *death*, the second accents *death* because hounding in itself is not fatal, and the third may be treated either way because figuratively there is a choice' (1972b: 639). Or again, he suggests that his theory of accent 'predicts that [ordinary meals] will carry no particular semantic weight, hence *Peter had CLAMS for dinner*, but that something in between [meals] may well do so, hence *I had some nice CLAMS for my SNACK this afternoon*' (Bolinger 1972b: 638).

In all the cases just cited (semantically empty nouns like *person* or *thing*; relative semantic weight of adjective and noun in phrases like *silky sheen* or *statistical approaches*; and sentence-final adverbs and prepositional phrases) the evidence for the highlighting view seems fairly compelling. However, cross-language comparison shows that the picture is not so clear. In languages like Italian, which resist deaccenting and which treat indefinite pronouns like any other noun phrase, virtually none of the English examples just cited has a counterpart with sentence stress patterns that reflect informativeness. For example, semantically empty nouns in Italian are treated for accentual purposes like any other noun:

(55) (a) ... perché ha ucciso un UOMO.
'... because he killed a man.'

(b) ... perché ha ucciso un POLIZIOTTO.
'... because he killed a policeman.'

In any context, *ha UCCISO un uomo* could only involve explicit contrast or metalinguistic correction (e.g. 'killed, not wounded'). Essentially the same is true of sentence-final adverbs and prepositional phrases. With these too we find very little tendency to deaccent and very little variability.

(56) (a) C'è una mosca nella MINESTRA.
'There's a fly in the soup.'

(b) L'hanno spaventato a MORTE.
'They scared him to death.'

As for analogues to English phrases like *silky sheen* or *statistical approaches*, it is not possible to construct these in Italian, because the normal word order in noun phrases is Noun+Adjective. However, it is worth noting that there is no tendency at all for Italian Noun+Adjective phrases to have the main accent on the noun, regardless of relative semantic weight.

6.2.2.4 Lexical and syntactic effects on relative semantic weight

It thus appears that there are languages in which relative informativeness is relevant to sentence stress and others in which it is not. However, those are not the only language-specific differences. Among languages that do take relative semantic weight into account, the details often differ unpredictably from language to language, or even from dialect to dialect within the same language. The conditioning factors are both lexical and syntactic.

A good illustration of the complexity of such cases is provided by English phrases consisting of a proper name and a common noun, such as *George Square* or *Alzheimer's disease*. As has been noted many times, when such phrases serve as names of thoroughfares (*Wellington Street*, *Chesley Drive*, *Dryden Road*, *Gillespie Crescent*, and so on), they have the main accent on the first part of the name if the second part is *Street* (i.e. *WELLINGTON Street*), but on the second part otherwise (i.e. *Chesley DRIVE*, *Dryden ROAD*, *Gillespie CRESCENT*). The explanation, proposed by an amateur observer of language well over half a century ago (see Mencken 1948) and repeated more than once since then, seems to be that in a town, *Street* is the least specific – and hence least informative – noun used in such names, and consequently can be deaccented.

In Ladd 1980b, I noted that the same sort of thing is true of several other sets of nouns, such as *Tompkins COUNTY*, *New York STATE*, *Baffin ISLAND*, but *GONDWANA land*. What I did not note is that there is consistent dialect variation in these cases. For example, in the set of nouns used in names of buildings, *House* is deaccented in American English (e.g. *FAUNCE House*, (Brown University Student Union building), *BLAIR House*, (Official US Government guest house in Washington); cf. *Morrill HALL*, *Carrie TOWER*, *Johnson MUSEUM*, etc.), but not in British English (e.g. *Adam HOUSE* (Edinburgh University examination hall), *Broadcasting HOUSE* (headquarters of BBC in London)). Nor did I note that definiteness can have an effect: when such phrases are lexicalised with the definite article, they are more likely to deaccent the common noun (e.g. *Rockefeller CENTER* in New York, but *the KENNEDY Center* in Washington). That is, while relative informativeness clearly plays a role, it does so within limits imposed by the lexicon and grammar of specific languages and dialects.

A good illustration of the interplay between relative informativeness and grammar is a common usage I observed while living in a rural area near Ithaca, New York. The official names of the rural roads almost all have *Road* rather than *Street* as the common noun (*Tunison Road*, *Townsendville Road*, *Lodi Center Road*, etc.), and as such should be accented on *Road* according to the rule of English just discussed. Yet in the rural context, the truly informative

element of the name is the proper name. Accordingly, when they refer to roads by name, local residents frequently add a definite article, which allows them to deaccent *Road* and put the main accent on the name (e.g. *the TUNISON Road, the TOWNSENDVILLE Road*). This example shows the importance both of relative informativeness and of specific grammatical and lexical effects. If, as claimed by the highlighting view, relative informativeness were the most important factor in sentence stress, then it should be possible for speakers simply to say *TUNISON Road* and *TOWNSENDVILLE Road*. But those pronunciations create unwanted narrow focus: adding the definite article seems to sanction the shift *while retaining broad focus*. It is difficult to explain this effect within the highlighting view.

A similar case is found in Icelandic. According to Kristján Árnason (2005: 453ff.), the word for 'man' may be deaccented in *definite* Adjective+Noun phrases and still convey broad focus, but not in indefinite phrases. Other more informative nouns bear the main accent in Adjective+Noun phrases regardless of whether the phrase is definite or indefinite. For example:

(57) (a) Þarna er GAMLI maðurinn.
'There's the old man.' (lit. there is old-DEF. man-DEF.)

(b) Þarna er gamall MAÐUR.
'There's an old man.' (lit. there is old man)

(c) Þarna er gamla PÓSTHÚSIÐ.
'There's the old post office.' (lit. there is old-DEF. post-office-DEF.)

What is important about these sentences is that they can all be used to convey broad focus, despite the fact that the main accent is on the noun in (57b) and (57c) and on the adjective in (57a). That is, Icelandic is like English in two respects: first, there are individual, relatively uninformative nouns that may be deaccented and still convey broad focus; and second, definiteness favours this kind of deaccenting where indefiniteness does not. But there is no direct correspondence between English and Icelandic: we cannot use *the OLD man* in English without conveying some sort of narrow focus on *old*.

Another similar case involves English and Hungarian phrases denoting sums of money. As we saw in example (4) above, in English, sums of money have the main accent on the unit of currency: *five FRANCS, fifty CENTS*, and so on. In Hungarian, however, it is normal to deaccent the unit of currency. Thus in contexts entirely comparable to ones discussed earlier, we might find:

(58) A: Mennyit kaptál érte?
'How much did you get for it?'

B: SZÁZ ötven forintot.
'150 forints.'

This comparison between English and Hungarian is very nearly identical to the hypothetical cross-language comparison sketched in section 6.1.3. In comparable contexts we have phrases containing corresponding words in the same order, yet we find one prominence pattern in one language and another pattern in the other. A Bolingerian or highlighting explanation for either pattern *on its own* is possible, but an explanation of the difference effectively destroys the predictive value of the highlighting theory. Universal strategies for highlighting salient information simply do not explain all the facts about accent across languages. The Hungarian pattern makes Bolingerian sense as an 'accent of interest' (the number of forints is what counts, and the unit is predictable), while the English pattern makes sense as an 'accent of power' (other things being equal, put the accent at the end). What does not make sense is that English treats the relative informativeness of the number and the currency unit as equal, while Hungarian treats them differently. There seems no way to avoid putting this kind of difference in the grammar of individual languages.

The Hungarian example raises the more general question of syntactic and phonological 'headedness'. In Hungarian the accentual pattern illustrated in (58) may actually reflect a general fact about noun phrases, not a specific fact about sums of money – that is, noun phrases normally tend to bear the main accent on the first element (viz. on the adjective in an adjective–noun sequence such as *nehéz nyelv* 'difficult language' in example (26) in chapter 2). Note that in the present example the number *száz ötven* itself is stressed on the first element; contrast English *a hundred and FIFTY*. It appears that some languages may generally prefer 'left-headed' phonological phrases – that is, phrases with the greatest prominence on the initial element rather than the final one. If this is true, then Hungarian is a likely instance of such a language: note that Hungarian word-stress is invariably on the first syllable of the word (cf. Halle and Vergnaud 1987, who suggest that prosodic headedness at the phrase level and the word level are instances of the same kind of structural specification at different levels in the prosodic hierarchy). The claim has even been advanced (Nespor and Vogel 1986: 168ff.) that phonological 'headedness' of this sort is systematically related to syntactic headedness, though the evidence for this claim is relatively thin (see further Ladd 2001: 1385f.).

One difficulty of evaluating any claims about 'left-headedness' – initial sentence stress – at the phrase level is due to factors of the sort discussed in section 6.2.1.3 above, namely to uncertainty about the validity of the phonological analysis, and specifically to the fact that in many such cases the question of constituent length is highly relevant. In the case of Hungarian noun phrases and number names, it seems likely that relatively short noun phrases do indeed have

the main accent on the first element, but that longer noun phrases (such as long number names) have a prominent accent on the last element as well. Extensive discussion of this general topic is given in Varga's excellent description of Hungarian phrasal stress patterns (Varga 2002: chs. 6 and 7). This effect of length is comparable to what I suggested about WHQs in Romanian in examples (25) and (26) above, and may be typical of 'left-headed' phonological constituents generally. Further examples are provided by the English intransitive sentences discussed in sections 6.2.3 and 7.2.3 below.

6.2.2.5 Summary

The English data reviewed in this section, in my opinion, represent the best evidence there is for the highlighting view. On the basis of the English cases alone, it is very clear that relative semantic weight is a key factor in sentence stress even in many broad focus cases. This fact is what drives the highlighting view, and is what induces investigators to try to understand more about the nature of 'relative semantic weight' (discourse status, informativeness, etc.): if we understand relative semantic weight, we will automatically understand sentence stress.

When seen in the light of cross-language comparison, however, the English data are less compelling. The totality of the evidence presented in this section suggests the rather less inspiring claim that relative semantic weight may play a role in sentence stress – in some syntactic contexts, in some languages. This does not make it any less important to study the contextual factors that influence relative semantic weight – for those languages in which it is relevant – but it does mean that in all languages there is an important role for structural considerations as well.

6.2.3 Predicates and arguments

We now turn to the third area of potential cross-linguistic differences in sentence stress patterns, namely the claim that in some languages there is a difference of accentability between arguments and predicates, with predicates (verbs and predicate nouns or adjectives) less accentable than arguments (noun phrases syntactically linked to a predicate). This claim, at least in the modern literature, seems to have been made first by Schmerling (1976: ch. 5). In origin it is a way of accounting for a long-standing and much-discussed problem with traditional normal stress rules.

In traditional syntax-based accounts of normal stress, the greatest prominence in a sentence is said to fall on the last content word. This rule encounters serious difficulties in English sentences with intransitive predicates. Some such

sentences seem to be most naturally pronounced with the main accent on the subject, but others seem more appropriate with the main accent on the predicate, and for others there is no agreement even among proponents of the idea of normal stress. There has been much discussion in the literature, over many decades, of the conditions under which the two accent patterns occur. These conditions are summarised briefly here.

In short sentences describing single events, nuclear accent on the subject is favoured:

(59) (a) The COFFEE machine broke.
(b) The SUN came out.
(c) His MOTHER died.

This is particularly true if the predicate denotes appearance or disappearance, or otherwise introduces the subject into the discourse (Allerton and Cruttenden 1979). By contrast, as was noted by Faber (1987), if the subject denotes a human agent, and the predicate denotes an action over which the subject is likely to have some control, nuclear accent on the verb is more likely. To use a term that has become rather more common than when Faber wrote, the cases in which the predicate is commonly deaccented generally involve *unaccusative* predicates (Perlmutter 1978; Levin and Rappaport-Hovav 1995; Sorace forthcoming):

(60) (a) My brothers are WRESTLING.
(b) Jesus WEPT.
(c) The professor SWORE.

Other intransitive sentence types that often bear accent on the predicate include sentences with generic subjects, and sentences that state definitions, eternal truths, and grand abstractions (Gussenhoven 1983a: 403ff.; Faber 1987: 352f.):

(61) (a) Wood FLOATS.
(b) Penguins SWIM.
(c) Hope FADED.

Accent on the verb is also more likely if the subject is (in some sense that is not very clear) 'topicalised', or otherwise readily referrable to the context: this accounts for the often-cited pair of examples in (62), first discussed by Schmerling 1976:41f.

(62) (a) Truman DIED.
(b) JOHNSON died.

These were the reports of the deaths of two former US presidents, addressed to Schmerling in both cases at the beginning of a conversation. Johnson's death was unexpected and the proposition (involving an unaccusative verb) is presented as a single event; by contrast, Truman's medical condition had been in the news for several days, and the sentence stress related the utterance to that prior context by treating *Truman* as a 'topic'. Similar subtly distinct pairs are discussed by Faber (1987) and Gussenhoven (1983a).

Given these data, Schmerling proposes that, in the absence of other considerations, arguments are intrinsically more likely to be accented than predicates. She treats cases where the main accent is on the intransitive predicate as special in some way – for example, what she calls 'topic–comment sentences' (1976: 89ff.), where the subject is somehow predictable or given in the context (as in (62a)). Bing (1979), Ladd (1980a), and Gussenhoven (1983a) all follow Schmerling by building a basic difference between arguments and predicates into their accentuation rules; and Gussenhoven follows Schmerling still further by discussing pragmatically different sentence types (e.g. his 'eventive' and 'definitional' sentences) which favour main accent on the subject or on the predicate.

Schmerling notes that the greater accentability of arguments also seems to hold in German: for example, object nouns are regularly more prominent than transitive verbs in main clauses (where the object follows the verb); in subordinate clauses (where the object precedes the verb); and in main clauses with non-finite forms of lexical verbs (where the object also precedes the non-finite verb). Thus:

(63) (a) Sie liest ein BUCH.
'She is reading a book.'
(b) Er denkt, dass sie ein BUCH liest.
'He thinks she is reading a book.'
(c) Sie hat ein BUCH gelesen.
'She read a book.' (lit. she has a book read)

The literal translation of the second example is *He thinks that she a book reads*; as Schmerling (1976: 84) points out, an English speaker reading such a literal translation aloud is likely to put the most prominent stress on *book*, not on *reads*. Selkirk (1984: sect. 5.2.2) makes similar observations about German, and Gussenhoven (1984: ch. 2) demonstrates at length that essentially the same is true of Dutch.

In all three languages, however, there is also a good case to be made that any apparent difference between nouns and verbs (or arguments and predicates) is

simply a matter of the more general role of relative semantic weight or informativeness. That is, it can be argued that nouns generally carry more semantic weight than verbs, which are frequently rather predictable given a particular noun or nouns. If this is true, then data like Schmerling's actually support the highlighting view, and any reference to structure (viz. to the predicate–argument distinction) is unnecessary. This case has been vigorously put by Bolinger (1972b; 1986: ch. 7; and especially 1989: ch. 9).

One type of evidence in favour of Bolinger's view is the existence of variability. There are many cases in which the main accent can plausibly occur on either a noun or a verb without necessarily signalling explicit contrast or narrow focus. For example, Bolinger points out that sentence stress may be variable in cases where 'nouns are set against verbs that are comparably low in semantic content' (1972b: 637):

(64) (a) I can't go with you; I've got too many THINGS to do.
(b) I can't go with you; I've got too many things to DO.

He notes that 'the same is true when noun and verb are equally rich' (1972b: 638):

(65) (a) We're looking for a neighbourhood where there are other boys to PLAY with.
(b) We're looking for a neighbourhood where there are other BOYS to play with.

(66) (a) It's too heavy a price to PAY.
(b) It's too heavy a PRICE to pay.

Another piece of evidence for Bolinger's view involves the effects of definiteness on the accentability of referring expressions (cf. section 6.2.2.4). Such cases have been discussed extensively with reference to German, where it is well established that definiteness affects both word order and sentence stress (e.g. Kiparsky 1966; Jacobs 1982). For example, in sentences that have a non-finite verb form at the end, the nuclear accent readily occurs on the verb if the preceding object noun phrase is definite, but not normally if the object is indefinite. The following examples are based on Cinque (1993):[11]

11 I have expanded Cinque's set of examples a bit to show that what is at issue is the *relative* acceptability of main accent on the verb, depending on the definiteness of the noun phrase. Cinque himself presents these examples as if the only possibilities were accent on the verb when the object is definite (67a) and accent on the object when it is indefinite (67d), but the situation is subtler than that. For more on the notion of relative acceptability and its theoretical implications, see Bard, Robertson, and Sorace 1996 and some of the references cited there.

(67) (a) Der Arzt wird einen PATIENTEN untersuchen.
'The doctor will examine a patient.'
(lit. the doctor will a patient examine)

(b) (?) Der Arzt wird einen Patienten UNTERSUCHEN.

(c) Der Arzt wird den PATIENTEN untersuchen.
'The doctor will examine the patient.'

(d) (OK) Der Arzt wird den Patienten UNTERSUCHEN.

When the object noun phrase is indefinite, as in (a) and (b), it is likely to refer to an entity newly introduced into the discourse, and placing the main accent on the verb is odd. By contrast, when the object is definite, as in (c) and (d), it is likely to refer to a 'given' entity, and is correspondingly less informative. As a result, it is much more acceptable to have the main accent on the verb.

Finally, whatever the semantics of the predicate or the information status of the arguments, sentence length can also play a role here, as with the sentence stress in WHQs (section 6.2.1.3 above; cf. section 7.3.1). Specifically, accenting of the subject in such intransitive sentences applies especially when the sentences are short: even with unaccusative predicates there may be accents on the predicate in sentences with longer constituents, and in sentences with additional adverbial constituents in the verb phrase.

(68) (a) Former President Johnson unexpectedly DIED today.

(b) The dog's mysteriously DISAPPEARED.

Here again, one might argue that a longer or 'heavier' constituent is more informative, and as such attracts the main accent; as already noted, we return to discuss such effects of constituent weight in section 7.2.3.

In all these cases, then, if we consider only data from West Germanic languages, it is once again almost impossible to decide between the highlighting view and an FTA account that assumes language-specific rules. However, if we compare extensively across languages, it becomes apparent that certain languages are like English in treating arguments differently from predicates in sentence stress rules, while other languages seem to treat arguments and predicates in the same way. Dutch and German, for example, are like English, while Italian and Spanish are not. It does not seem useful to account for the English or German data solely in terms of universal principles of informativeness.

Direct comparison of English or German with Spanish or Italian is unfortunately complicated by the fact that it is difficult to find analogues to sentences like *The COFFEE machine broke* (59). In general, such intransitive event sentences in Spanish or Italian have verb–subject (VS) word order:

(69) S'è rotta la CAFFETTIERA.
'The coffee machine broke.'
(lit. has broken the coffee machine)

That is, as with the difference between negative and non-negative indefinite pronouns (section 6.2.2.2), word-order modifications in languages like Spanish or Italian may indirectly achieve the accentual effects that English accomplishes directly by manipulating the location of the nuclear accent. In fact, VS word order in Spanish or Italian occurs in circumstances that are uncannily parallel to those in which intransitive predicates are unaccented in English, and in any given context, the nuances of VS versus SV word order in Spanish or Italian are generally extremely similar to the nuances of unaccented versus accented intransitive predicates in English (for examples of this see Bolinger 1954; Hatcher 1956; Ortiz-Lira 1994). This observation gives some comfort to a highlighting view of sentence stress, because it means that 'accent on the argument' or 'accent on the predicate' appear to have similar pragmatic effects in different languages. At the same time, however, the highlighting account of these similarities provides no explanation for the fact that some languages (like English) can readily move accent away from the last content word, while others (like Italian) must modify word order to get the appropriate content word into accented position.

More direct evidence of a difference between English or German and Italian or Spanish with regard to the relative accentability of nouns and verbs comes from certain specific constructions. These include infinitive 'small clauses', such as (70a), and short relative clauses containing no nouns, such as (70b)

(70) (a) I have a BOOK to read.

(b) It was caused by the FISH she ate.

In English, these often have the main accent on the last noun rather than on the following verb (cf. examples 64–66). In Italian, corresponding sentences are accented on the verb:

(71) (a) Ho un libro da LEGGERE.
'I have a book to read.'

(b) È stato provocato dal pesce che aveva MANGIATO.
'It was caused by fish she had eaten.'

It is true, as Bolinger points out, that these kinds of structures are variable in English in phrases like *things to do* and *price to pay*, so it may be inappropriate to take the pattern in (70) as revealing the neutral sentence stress pattern in English. Nevertheless, there is clearly a cross-language difference, because

this variability is absent in Italian: that is, Italian phrases like *cose da fare* ('things to do') or *prezzo da pagare* ('price to pay') can only be accented on the verb – except, as always, in cases of very explicit metalinguistic contrast. Unlike English, Italian really does put the main accent on the rightmost content word in the overwhelming majority of contexts. This, in turn, means that the apparent difference of accentability between predicates and arguments in English is not simply a specific instance of more general principles of accentual highlighting, but involves a language-specific, structure-based rule of sentence stress.

The examples just cited, in addition to whatever they may tell us directly about sentence stress, also make us aware of a methodological consideration: if we are looking for unambiguous differences between languages in the relative accentability of nouns and verbs, we need to look at structures in which the verb is sentence-final. Such structures are difficult to find in comparing English and Italian. A more obvious place to look is in languages with subject–object–verb (SOV) word order. Specifically, one would predict that in an SOV language in which arguments are more accentable than predicates, the neutral or broad focus location for the main accent would be on the object, whereas in an SOV language that does not distinguish predicates and arguments accentually, it would be on the verb.

There does appear to be such a difference between SOV languages. In Turkish, especially with indefinite objects, the last accent normally falls on the object (see footnote 7 on p. 225):

(72) Eski müdür bir KITAP yazdı.
'The former director wrote a book.'
(lit. former director one BOOK wrote)

In Bengali, on the other hand, the last accent is generally on the verb (Hayes and Lahiri 1991):

(73) Ram Shamoli DEKHLO.
'Ram saw Shamoli.'
(lit. Ram Shamoli saw)

This difference would appear to confirm the existence of a difference between languages that treat arguments and predicates unequally and languages that do not. However, once again we are bedevilled by disagreements over the data. In a typological survey article by Kim (1988, cited in Cinque 1993), the claim is put forth that most SOV languages, including Bengali, have unmarked sentence stress on the object. This directly contradicts the data from Hayes and Lahiri just cited. While we might grant prima facie credibility to Hayes and Lahiri on the grounds that Lahiri is a native speaker of Bengali, it is clear that this is a point on which further data are required. Broad focus sentence stress patterns in languages with SOV word order need to be scrutinised.

6.3 Typology of sentence stress patterns

The existence of cross-linguistic differences of sentence stress leads naturally to the question of typology, and of constraints on the ways in which sentence stress patterns can differ across languages. Can languages differ 'without limit and in unpredictable ways' (to use the often-quoted formulation of Joos 1957: 96)? Or is the range of variation constrained in principled ways? Or, for that matter, is the apparent variation all conditioned by some other property of a given language?

The first possibility – that there are no principled limits on the way sentence stress can differ from language to language – cannot be logically excluded, but it runs counter to much current thinking on language typology, as well as contradicting many long-standing observations about universal tendencies in intonation, and I do not pursue it further. The third possibility – that the apparent cross-linguistic diversity of sentence stress is actually a reflection of something else – is suggested in various works broadly in the tradition of generative syntax. For example, Cinque (1993) argues for a 'null theory of sentence stress', in which the most prominent accent of the sentence goes on the most deeply embedded element in the sentence's surface syntactic structure, and therefore claims that apparent differences in sentence stress are really syntactic differences. A similar position is extensively developed by Zubizarreta (1998), though on the basis of a different version of generative syntax. Irrespective of one's preferred approach to syntax, I do not believe that this view can be maintained in light of the data presented in the foregoing subsections; however, a point-by-point rebuttal of Cinque's or Zubizarreta's arguments is beyond the scope of this chapter.

This leaves us to search for principles that constrain the variability of sentence stress patterns from one language to another. Given the discussion so far, it seems clear that in one way or another the principles involve some sort of balancing act or competition between different principles – in a given sentence, one principle (say, definiteness) may favour placing the main accent on word *x*, while another competing principle (say, relative accentability of arguments and predicates) may favour placing it on word *y*. This, in turn, suggests that a description expressed in terms of Optimality Theory (Kager 1999; Prince and Smolensky 2004) would be attractive, and indeed such descriptions have been discussed, for example by Féry and Samek-Lodovici (2006) and German,

Pierrehumbert, and Kaufmann (2006). I do not attempt to work out such an analysis here, but in this section I briefly identify some of the potentially competing principles that must be involved.

First of all, it seems fairly clear that in all languages, even in a language as strongly 'left-headed' as Hungarian, there is some tendency for the main accent to occur near the end of a phrase or sentence. This is Bolinger's notion of 'accents of power'. This principle may actually be a composite of two more basic principles, namely a tendency to avoid having the main accent early in the sentence, and a tendency to put the main accent on content words. Given this 'rightmost' tendency, however, there appears to be a fairly sharp division between languages in which rightmost main accent is overwhelmingly the norm (except in cases of explicit metalinguistic correction), and languages that allow the main accent to be placed earlier in the sentence for a variety of other reasons (including Bolinger's 'accents of interest'). In the former type, the neutral location of the main accent is rightmost in all or nearly all cases, and none of the factors we have been discussing – question versus statement, relative informativeness, part of speech, definiteness, phonological headedness, and so on – plays any important role. These are Vallduví's [–plastic] languages, such as Italian and Catalan. In the latter type – Vallduví's [+plastic] languages such as English or Greek – the factors we have been discussing permit the main accent to occur further to the left. In Optimality Theory terms, the constraint by which the main accent must be rightmost would dominate most or all other sentence stress constraints in, say, Catalan, but would be dominated by other constraints related to informativeness in English, by constraints relating to question focus in Greek, by constraints related to phonological headedness in Hungarian, and so on.

In languages where the main accent is not exclusively on the rightmost content word, another obvious typological question is whether there is any systematic relationship among the various factors that draw accent to the left. For example, given the apparent parallel between YNQs and WHQs, it is tempting to try to relate their behaviour – and indeed, this is what I have implicitly done by treating both types in the same section. That is, it may be that some languages (e.g. Romanian and Hungarian) treat questions and statements differently for accentual purposes, while others (e.g. English and Italian) treat them alike. Alternatively, it may be that the accentual behaviour of YNQs and WHQs in any given language is independent: there may be languages that place the main accent on the verb in YNQs but treat WHQs like statements, and languages that place the main accent on the WH-word in WHQs but treat YNQs like statements. Superficially, it seems straightforward to determine whether such cases exist, but there are plenty of complications,

such as the fact that WHQs in Bengali require a narrow focus tune (Hayes and Lahiri 1991: sect. 4.2); or the fact that focus in YNQs in Turkish is marked in the first instance by the question particle and only secondarily by sentence stress; or even the simple fact that in some languages WH-words appear at the beginning of the clause and in others they occupy positions ordinarily filled by argument noun phrases. Furthermore, there is not much reliable information about sentence stress in questions in many languages; typological research in this area lacks a broad empirical foundation.

Similarly, we might speculate that all the factors relating to accentability and informativeness – contextual deaccenting, the special treatment of indefinite pronouns and 'empty' nouns, and effects of definiteness – are systematically related to the differential treatment of arguments and predicates. For example, we might speculate that if a language treats predicates as less likely to be accented than arguments, it will also tend to deaccent contextually given material and exhibit variability in sentences involving indefinite pronouns or empty nouns. The basis for this speculation is Bolinger's suggestion that the argument–predicate difference is a manifestation of the general relationship between accentuation and informativeness. As with accentuation in questions, it would seem straightforward to determine whether these dimensions of variation are independent, but once again there are complications, ranging from the effect of sentence length on accentuation to the existence of different word-order patterns for predicates and arguments in different languages of the world. It therefore seems useful to re-emphasise how little we know about accentuation cross-linguistically. I believe that the exemplification of cross-linguistic differences given here, combined with the general expectation that typological variation should be highly constrained, provides a clear focus for empirical investigation.

7 *Phonological issues in sentence stress*

Evidence of the kind discussed in chapter 6, in my opinion, makes any universal highlighting view of sentence stress untenable. However, there are obvious theoretical vulnerabilities in the FTA theory as well, which is one of the reasons that broad focus lives on as a topic of empirical study and theoretical discord. Some of these problems are essentially phonological, and revolve around the fact that we used the same notation *five FRANCS* to indicate both broad focus and narrow focus on *francs*. This notational practice implies two claims about sentence stress that require justification. First, it implies that some sentence stress patterns are *ambiguous* between different focus interpretations; second, it implies that a phrase or utterance must contain a single most prominent (primary) accent. These specific topics, and their broader implications for the phonology of sentence stress, are the subject of this chapter.

7.1 Phonological problems with the Focus-to-Accent view

7.1.1 Is the broad focus pattern really ambiguous?

In this section we scrutinise the claim that the broad focus sentence stress pattern *five FRANCS* can be identified with one of the narrow focus patterns, and that the pattern notated this way is therefore potentially ambiguous. If broad focus and narrow focus readings can always be distinguished, then broad focus (or focus projection) presents no fundamental difficulty for the highlighting view; it requires only a more elaborate account of the broad focus pattern as a whole. There is certainly no doubt that it is *possible* to pronounce the phrase *five francs* in such a way that it fairly unambiguously conveys narrow focus on *francs*. Yet a variety of evidence suggests that some patterns of sentence stress really are ambiguous.

One line of evidence comes from the results of experimental studies of the perception of accent patterns. Gussenhoven (1983b) showed that, while English listeners can distinguish sentence stress patterns analogous to the broad focus

and contrastive stress version of *five FRANCS* under certain conditions, they are still much more likely to confuse these patterns with each other than they are to confuse them with patterns analogous to *FIVE francs*. A much larger study by Rump and Collier (1996) shows similar effects for Dutch. Admittedly, this evidence can be looked at in two different ways. Rump and Collier, for example, emphasise the potential distinctiveness of the corresponding forms in their data, rather than their confusability, but their data show clearly that the pattern corresponding to *FIVE francs* is acoustically quite distinct from the other patterns.

Clearer evidence for a genuine linguistic ambiguity between broad and narrow focus comes from jokes and wordplay. One example is the exchange that is supposed to have occurred between a reporter and Willie Sutton, a notorious American bank robber of the 1930s and 1940s, shortly after Sutton's arrest:

(1) Reporter: Why do you rob banks?
Sutton: Because that's where the money is.

The reporter's question might be said to have broad focus on the entire verb-plus-object phrase *rob banks*; that is, the point of the question is why Sutton engages in the antisocial activity of robbing banks. Sutton's reply – which clearly works as a wisecrack or witty repartee – treats the question as if it focuses only on *banks*, that is, as if the questioner presupposes that Sutton engages in robbing something, and wants to know why he robs banks rather than, say, grocery stores or filling stations.[1] Essentially the same joke is found in a *Peanuts* cartoon (7 February 1994). Charlie Brown, in bed, says, 'Sometimes I lie awake at night, and I ask, "Why am I here?"' In the next panel, he continues, 'Then a voice answers, "Why? Where do you want to be?"' Charlie Brown's question is intended as a broad focus existential question that might be paraphrased as 'Why do I exist?', but the voice that answers treats the accent on *here* as if it signals narrow focus – as if Charlie Brown is asking why he is 'here' rather than somewhere else.

One reasonable way to make sense of this evidence is to say that the two patterns that we have been writing as *five FRANCS* – viz. the broad focus reading and the reading with narrow focus on *francs* – are in fact linguistically identical, as the notation implies, but that the two readings can be distinguished through the use of additional phonetic *emphasis*. Emphasis, in this view, would be a

1 The authenticity of the Willie Sutton quote is doubtful. More information about the history of this anecdote can be obtained by simply entering the complete phrase *that's where the money is* into any search engine.

paralinguistic device that can sometimes be brought into play: that is, we have the linguistic distinction between two categorically different accent patterns, *FIVE francs* and *five FRANCS*; and in addition we have the paralinguistic possibility of gradiently modifying the realisation of those patterns so as to single out individual words. 'Emphasis' and 'accent' may often go hand in hand, but that does not mean they are the same thing.

This point of view – which is based on the discussion of gradience in section 4.2.2 – can be made more plausible if we see the location of sentence stress as analogous to the location of focus particles. In many cases, in many languages, focus particles have a neutral or default location in the sentence, but can be attached in other locations to convey narrow focus. If the particle is in the neutral location, there is thus often a potential ambiguity between a broad and a narrow focus reading. In these cases, as with focus conveyed by accent, different degrees of local emphasis may help clarify the intended meaning. For example, in Russian the negative particle *ne* occurs before a narrowly focused constituent if there is one, but otherwise precedes the finite verb. If narrow focus is intended on the verb, this can be made clear with emphatic accent (indicated by italic in the following example):

(2) a) Ne MARK čitaet.
'*Mark* isn't reading [someone else is].' (lit. not Mark reads)

b) Mark ne ČITAET.
'Mark isn't reading.'

c) Mark ne *ČITAET*.
'Mark isn't *reading* [he's doing something else].'

That is, the interpretation of the focus as broad or narrow is made clear by the presence or absence of local emphasis, not by the presence of accent or the presence of the negative particle alone.

When focus is conveyed by particles rather than accents, it is uncontroversial that the choice of focus pattern is categorical: either the focus marker is attached to one word or it is attached to another. It is also uncontroversial to distinguish the linguistic feature (the location of the focus particle) from the paralinguistic feature ('emphatic stress' to signal narrow focus), because the linguistic feature is segmental and the paralinguistic feature suprasegmental. The proposal here is that in languages like English, linguistic accent location and paralinguistic degree of emphasis – even though they are both phonetically suprasegmental – should be distinguished in the same way. This raises once again the issues discussed in section 4.2.2, concerning 'gradience' and intonational meaning, but the general type of account I am proposing is that there are certain accent

patterns whose interpretation is ambiguous between broad and narrow focus, and which can be disambiguated by paralinguistic means.

All this evidence points to the conclusion that there are only two basic sentence stress patterns in a phrase like *five francs*: one with the sentence stress on *five* and one with the sentence stress on *francs*. (I am here ignoring the 'double-focus' cases illustrated in example (3) in chapter 6, and will return to them in the next chapter.) The accent on either word can be exaggerated or made phonetically more prominent, by increasing pitch range, intensity, and so on, in order to emphasise the individual accented word. This emphasis helps to make clear that the accent signals a narrow focus: if the accent on *five* is exaggerated in this way, the emphasis is redundant, but if the accent on *francs* is exaggerated, the emphasis helps to disambiguate the focus interpretation. But in either case, the emphasis can be seen as a modification of the two basic patterns.

7.1.2 *Is there a single 'nuclear' accent?*

Another important problem for the FTA view is that the notion of broad focus is based on the location of a *single* accent in each sentence or phrase – usually the last accented word. This practice is obviously informal, and has been called into question by a few authors (notably Bing 1980 and Bolinger 1986), but it is surprisingly common throughout the literature, even among proponents of the highlighting view, and it is not hard to show that there is empirical justification for it. It is therefore useful to try to put it on a more secure theoretical footing.

It is, first of all, fairly clear that there is no *phonetic* justification for treating the last accent in a sentence as more prominent. The last accent is often low in pitch and in overall intensity, and Bolinger treats the intuition that it is more prominent as nothing but an illusion, somehow induced by its position in the utterance (see Bolinger 1986: 58ff. and ch. 6 *passim*). (He does not attempt to explore the possibility that the existence of such an illusion might be interesting or revealing.) The IPO tradition identifies the prominence of an accent in explicitly phonetic terms, relating it to the size of the pitch excursion (e.g. 't Hart, Collier, and Cohen 1990; Terken 1991; Hermes and Rump 1994; Rump and Collier 1996), and allows no theoretical reason for regarding the final accent as anything special. Nevertheless, it is easy to show that in most contexts the last accent does have a special status in signalling focus: the location of the last accent determines whether a broad focus interpretation is possible. This can be illustrated by taking a short phrase and gradually expanding it.

Consider a phrase like *a cup of coffee*. The basic facts about sentence stress and focus in this phrase are completely comparable to those for the example

five francs treated above. Broad focus can be conveyed by the pattern *a cup of COFFEE*, while only a narrow focus reading is available for the pattern *a CUP of coffee*: that is, the first pattern is potentially ambiguous between focus on the whole phrase (a cup of coffee as opposed to a sandwich or a piece of cake) and focus on *coffee* (a cup of coffee rather than a cup of tea or cocoa); the second unambiguously focuses on *cup* (i.e. a cup of coffee rather than a pot of coffee). As with the example *five francs*, it is not necessary for the accent on *coffee* to be the only accent of the phrase in the broad focus pattern: *cup* may be accented as well, without affecting the focus. But it is necessary for the accent on *coffee* to be the *last* accent of the phrase, and that is what is expressed in the notation *a cup of COFFEE*.

The importance of the location of the last accent can be seen by expanding the phrase, to something like *five francs and a cup of coffee*. A broad focus reading of this phrase would be appropriate in the context we considered in example (5) in chapter 6:

(3) A: What did they give you for participating in the experiment?
B: Five francs and a cup of coffee.

In order to signal broad focus, there *must* be an accent on *coffee*. If the last accent is anywhere else (e.g. *five FRANCS and a cup of coffee* or *five francs and a CUP of coffee*), some sort of narrow focus is implied. As has often been observed, the number of accents may depend on the deliberateness of the speech and a number of other factors. If speaker B is answering hastily and matter-of-factly, it may be difficult to distinguish any accent other than the one on *coffee*, and perhaps an H* accent on *five*. If Speaker A had already asked the question and been given the answer earlier, Speaker B might say impatiently:

(4) L* L* L* H* LL%
FIVE FRANCS and a CUP of COFFEE. [I already told you that!]

If speakers A and B are fellow students in a department that requires them to participate in experiments for minimal reward, B might say:

(5) H* !H* !H* !H* LL%
FIVE FRANCS and a CUP of COFFEE. [... the usual].

But none of these differences matters for the focus interpretation: all that counts is that there is an accent on *coffee*.

We can extend the example even further – perhaps stretching the bounds of plausibility a little – to something like

(6) Five francs seventy-five centimes and a cup of pretty tasteless coffee.

Again, in order for this phrase to serve as a broad focus response to the question about what the speaker was given for participating in the experiment, the last accent must be on *coffee*. As in the previous example, the overall number of accents and the details of any expressive intonational choices are irrelevant; but as in the previous example, any location of the last accent other than on *coffee* signals some fairly specific contextual presuppositions, and narrow focus.

The special significance of the last accent, then, lies not in any actual phonetic prominence it may have, but in the key role it plays in defining the *pattern* of prominence. It need not be specially prominent; it need only be present. This explains the emphasis on describing the location of the rightmost accent in much previous work on sentence stress. More generally, it justifies our earlier suggestion (section 4.1.1) that the nuclear accent has some special structural status, and a special role in signalling focus. We return to this point repeatedly in what follows.

7.1.3 *On the definition of 'pitch accent'*

Finally, let us briefly address a fundamental problem with the FTA view which attracts less critical attention than it should, namely the definition of pitch accent. This is the point at which 'pitch' and 'prominence' intersect, and we still do not really understand the relation between the two aspects. The definition given in section 2.2 ('a local feature of a pitch contour – usually but not invariably a *pitch change*, and often involving a local maximum or minimum – which signals that the syllable with which it is associated is *prominent* in the utterance') is crucially vague about both pitch and prominence, and ultimately leaves a great deal to the judgement of the analyst. In the online ToBI training materials, the notion of pitch accent is extensively exemplified but never really defined; the distinctions between examples are expressed in terms of different 'levels of stress' and very much depend on the observer's impressions of what sounds more 'prominent' or more 'salient'. I do believe that this reliance on intuitive judgements of relative prominence in an actual utterance context is appropriate, but it has theoretical implications that have not yet really been explored.

In most cases, transcribers are reasonably reliable and consistent in their judgements about which words are accented. The main reliability studies on ToBI transcription show that inter-transcriber agreement on the location of pitch accents is very high: for English, Pitrelli, Beckman, and Hirschberg 1994 found 81 per cent agreement and Syrdal and McGory 2000 found 91 per cent, while for German, Grice *et al.* 1996 found 87 per cent agreement. However, there are at least two types of situations in which disagreements are

not uncommon. One of these is the case of postnuclear accents, which we have already discussed in section 4.1.4, and which we will return to in section 8.1.2. The other involves sequences of prenuclear accents.

The difficulty with prenuclear accents comes from the fact that they are (as we saw in section 4.1.1) less prominent in some way than the nuclear accent that follows them. In a short phrase with only two prominent words, this creates few problems. We have already seen several examples of such phrases, and in fact in section 2.4 we were able to use intuitions of relative prominence to motivate the distinction between downstepped and non-downstepped accents in English. The relevant examples (20–22) from chapter 2 are repeated here for convenience, with an AM representation of the intonation patterns:

(7) H* L L%
(a) my MOTHER'S diaries [strong–weak, focus on *mother*]
H* H* LL%
(b) my mother's DIARIES [weak–strong, non-downstepped]
H* !H* LL%
(c) my mother's DIARIES [weak–strong, downstepped]

The relative prominence of the two accents is intuitively easy to judge because of the differences of pragmatic interpretation that result: only in the case where *mother's* is more prominent than *diaries* do we detect an element of strong pragmatic contrast between *mother* and some other possible owner of the diaries.

However, when we have three or more prominent words in a short phrase, especially a phrase involving downstep, the status of the middle word or words is often unclear. This can easily be illustrated if we expand the phrases in (7) to include a third content word. For example:

(8) (a) My wife's mother's diaries
(b) My late mother's diaries
(c) My mother's college diaries
(d) My mother's old diaries

These can all be pronounced with distinct local pitch peaks on all three content words, in which case most transcribers would unhesitatingly indicate three H* pitch accents, with the nuclear accent always on *diaries*. However, they can also all be pronounced with a high initial pitch peak on the first content words,

and a steadily declining pitch across the rest of the utterance, but with *diaries* still clearly the (downstepped) nucleus, as in (7c). We will therefore have an H* on the first word and an !H* on *diaries*, but what of the middle word? Do we regard it as having an !H* accent, or no accent at all? Transcribers' intuitions are much less clear on this point.

The original ToBI training materials contain a number of such sentences, with markedly varied 'official' transcriptions of the medial prominent syllables. Some examples are shown in (9). In these examples, the syllables that I judge as impressionistically prominent are shown in bold, while the tonal transcription is the one provided in the training materials. (Names in square brackets following the sentences are the ToBI filenames.) It can be seen that some of the medial syllables shown as impressionistically prominent are transcribed with an accent involving a downstepped H (9a and 9b); some are transcribed only with /*?/, indicating uncertainty about presence of pitch accent (9b and 9c); and some are transcribed with no indication of pitch accent at all (9d and 9e).

(9) (a) H* !H* !H* L-L%
A **friend** of mine **works** for **NA**SA ['friend2']
(b) H* !H* ?* !H* L-L%
He **sold** the **bus**iness to **some**body **else**. ['sold1']
(c) H* *? *? !H L-L%
...**State Supreme Court opin**ions. [part of 'author']
(d) L+H* !H* L-
Capote **died Sa**turday [part of 'capote']
(e) H* L+!H* L-L%
[No] I **think** I'll **wear** my **hik**ing boots. ['wellies2']

I do not mean to suggest that the transcriptions are wrong – though I certainly disagree with some of them. Although the overall patterns are quite similar, there are some clearly audible distinctions between the medial syllables of the various sentences, and these may be the basis of some of the differences between different transcriptions. But one important source of the uncertainty can be seen clearly in figure 7.1, which shows the F_0 trace for example (9a). Here, and in the other cases in (9), we are dealing with syllables that form part of a more or less smoothly descending pitch contour, and in this intonational context the judgement that a word is accented cannot be based on any obvious pitch change or pitch excursion. Instead, transcribers must be basing their decision on other phonetic cues such as duration.

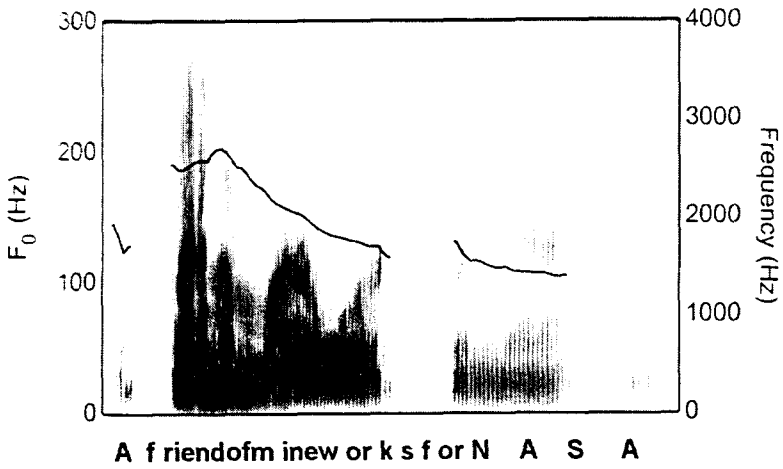
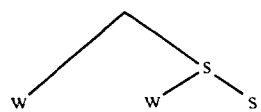


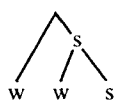
Figure 7.1 *The English sentence* A friend of mine works for NASA, *spoken as a declarative sentence in a larger narrative in which it provided crucial background knowledge for what followed. In the ToBI training materials this is transcribed with accents on* friend, works, *and the first syllable of* NASA (['næsə]), *in a downstepping accent sequence H* ... !H* ... !H*_L_L%. The pitch contour declines smoothly from the initial peak, with no local pitch movements corresponding to the perceived pitch accents on* works *and* NASA. *Note that the F_0 extraction failed on the very weak final syllable of* NASA.

A possible metrical account of the uncertainty goes like this. In at least some of the cases we have a right-branching metrical structure like the following:

(10)



(a) A friend of mine works for NASA.



(b) Capote died Saturday.

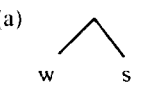
On one interpretation of metrical trees (Giegerich 1985: 18ff.), this structure results in an undefined prominence relation between the two weak branches of the tree: both are subordinate to the strong branch, but neither is subordinate to the other. One possible transcription (the official transcription of (9a)) treats this lack of defined prominence relation as indicating a two-level prominence distinction (i.e. *NASA* > *friend* ≡ *works*), by putting prenuclear accents on both weak branches. An alternative possibility (like the official transcription of (9d)), in the absence of defined relative metrical prominence, is based instead on some intuitive notion of relative acoustic prominence: the first weak branch is more acoustically prominent than the second because of the initial pitch peak at the beginning of the downstepping contour, so the transcription implicitly encodes a three-level prominence distinction (i.e. *Saturday* > *Capote* > *died*) by putting a prenuclear accent on *Capote* and no accent at all on *died*. Ultimately, such cases will only be resolved when we better understand the balance between acoustic and metrical factors in transcribers' judgements of relative prominence.

7.2 A metrical theory of sentence stress

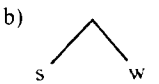
In the preceding section I discussed the sense in which the last accent in a phrase or sentence may be regarded as prominent simply by virtue of its position in structure. I also drew attention to problems with the notion of pitch accent – problems that seem to point to the need to recognise position in structure as an explicit part of the definition of prominence. In this section I present evidence for the more general claim that the link between focus and sentence stress is ultimately mediated by a hierarchical prosodic structure. Specifically, I aim to show that it is not pitch accent per se, but *relative metrical strength* that is the essential signal of focus. In many cases, of course, metrical strength is manifested as pitch accent, but in a number of cases that I will discuss, it is necessary to recognise the priority of metrical strength. This general view is consistent with the discussion in the preceding section, and with the overall attempt in this chapter and the next to integrate metrical and autosegmental aspects of prosodic description.

For simplicity, I will retain the capital-letter notation to indicate sentence stress, but will give it an explicitly metrical interpretation: the capitalisation of a word can be taken as a shorthand indicating that the word is the *Designated Terminal Element* (DTE; see section 2.2.2) of the utterance or of some relevant subconstituent. That is, the notations *five FRANCS* and *FIVE francs* are equivalent to the metrical notations in (11a) and (11b) respectively:

(11) (a)



(b)



This expresses the idea that sentence stress above all involves a syntagmatic relation between two constituents.

7.2 *'Stress-first' and 'accent-first' theories*

The difference between the FTA view discussed in chapter 6 and the view I wish to develop here is the difference between what Selkirk (1984: sect. 5.3) called 'accent-first' and 'stress-first' theories of sentence stress. Selkirk's own account of the focus–accent relation is what she calls an 'accent-first' account. In her model, a 'metrical grid' is constructed for each utterance, representing various features of the utterance's rhythm and stress pattern. Independently, pitch accents are assigned to individual words. The assignment of pitch accents pays no attention to the metrical grid, but is based purely on syntactic–semantic constituents ('sense groups'), in accordance with the Focus-to-Accent theory. In some cases this results in pitch accents being placed on words that do not have maximum (level 5) prominence in the metrical grid, in which case the grid must be adjusted so that pitch accents are only on maximally stressed syllables. (Among other things, this means that all pitch accents are, in theory, equally prominent.) Apart from this adjustment of the metrical structure, there is no connection between pitch accent and stress, and Selkirk is at pains to insist that the string of pitch accents should not be seen as an additional level in the grid. She contrasts her approach with a 'stress-first' view, in which focus would be reflected directly in the metrical grid; the pattern of prominence in the metrical grid would, in turn, control the distribution of pitch accents in some way.

When Selkirk wrote, the question of whether stress or accent 'comes first' could be interpreted literally, within the context of a derivational model of phonology: pitch accents (or metrical structure) are assigned, 'and then' something else happens in the derivation. But this derivational interpretation – as Selkirk herself makes clear – is not essential. The dichotomy between accent-first and stress-first can still be understood in other ways that are more compatible with other ideas of what phonology is like. Stripped of its derivational overtones, the question boils down to something like the following. Assume that there is a phenomenon of postlexical or utterance-level prominence whose principal function is to signal focus. Is this postlexical prominence primarily a question of the distribution of pitch accents ('accent first'), or primarily a matter of relative prominence in the phonological structure of the utterance ('stress first')?

The evidence strongly favours the stress-first view. There are two kinds of cases that ought to be fatal counterexamples for any accent-first version of the FTA theory, namely accents that do not signal focus ('accent without focus'), and focus that is not signalled by accent ('focus without accent'). Both kinds of cases occur.

7.2.1.1 Accent without focus

The case of accent without focus regularly arises in broad focus renditions of phrases and sentences. Although, as we saw in section 7.1, there is a good deal of uncertainty and disagreement over the phonetic details of broad focus, and even the definition of pitch accent, it seems clear that detectable phonetic prominence can occur on words preceding the nuclear accent. Given strict accent-first assumptions, the existence of such prominence is a source of some theoretical consternation: if in a phrase like *five FRANCS* there is no separate focus on *five*, then there should be no accent either.

In some contexts it is possible to ignore the problem for phonetic reasons: on a phrase-initial monosyllabic word such as *five*, any phonetic prominence may be difficult to detect. However, if we replace the phrase *five francs* by *my mother's diaries*, the presence of an accent of some sort on *mother* is more or less indisputable. In the terminology proposed by Bolinger, the accent on *mother* would be called a 'B accent', while that on *diaries* would be an 'A accent'. In Pierrehumbert's terms, both accents would be H*, but the second would be followed by an L phrase accent and the first would not. Whichever taxonomy we adopt, the prenuclear accent is clearly there, causing problems for a strict interpretation of the FTA view.

Selkirk (1984: 274) discusses just this point in connection with single-word utterances like *California*. Since this word constitutes the focus in a single-word utterance, it should have a pitch accent, and indeed it does. Unfortunately, it frequently has *two* pitch accents – one on the primary stressed syllable *-for-* and one on the secondary stressed syllable *Cal-*. Selkirk admits that she has no explanation for the 'additional' accent. There are similar problems with Gussenhoven's Sentence Accent Assignment Rule (e.g. 1983a), which assigns a single main accent to each 'focus domain', irrespective of the number of content words it contains; like Selkirk, Gussenhoven is aware of the problem of additional accents (1984: 89, n. 4), but does not deal with it in detail. Steedman's early work (e.g. Steedman 1991; Prevost and Steedman 1994) has more serious problems with these additional accents, because his analysis is explicitly predicated on the presumed *absence* of accent on any word that is not the head of a focus constituent.

Enough evidence of such problems has accumulated that many authors (e.g. Kruyt 1985; Welby 2003; Calhoun 2006) now explicitly recognise a distinction between primary and secondary (or nuclear and prenuclear) accents. Given such

a distinction, we could state that only primary or nuclear accents are relevant to signalling focus, and that secondary accents are distributed according to other criteria. This would make it possible to retain the basic ideas of an 'accent-first' account of sentence stress within the general framework of the FTA theory; specifically, we lose little of substance if we abandon the insistence on equal pitch accents, and assume some kind of difference between nuclear and other accents. Selkirk (1984: 274) actually anticipates some such solution in her discussion of the *California* example just cited:

> The additional pitch accents do not seem to modify the focus-related properties of the utterances ... Whether additional pitch accents like these have the same status in the grammar, and in particular whether they are assigned in (surface) syntactic structure and associated by [Selkirk's Pitch Accent Association Rule], we will leave open. [Selkirk's theory of the intonation-stress relation] does not require us to ascertain the status of these secondary pitch accents.

This general solution to 'accent-without-focus' thus appears remarkably compatible with much past work, and, as we shall see in chapter 8, seems likely to form part of any viable theory of sentence stress.

7.2.1.2 Focus without accent

Even if we recognise a distinction between primary and secondary accents, however, problems still remain. The most obvious is the occurrence of 'focus-without-accent' – cases where focus is signalled by phonetic cues to prominence such as duration and vowel quality, apparently without any pitch accent whatsoever. Many such cases in English involve prepositional phrases with pronoun or adverb objects, such as *for him* and *in there*. These can of course occur with pitch accent on either word, but exactly the same focus distinctions can be made when the entire phrase is completely deaccented. This is, as far as I can see, impossible to reconcile with any version of the accent-first FTA view.

Consider the phrase *for him*. This can be pronounced with either word bearing the main accent:

(12) (a) I did it for HIM.
(b) I did it FOR him.

Main accent on *him* clearly conveys narrow focus – 'for him and not for someone else'. Main accent on *for* apparently conveys broad focus, for example 'I did it for him because he wouldn't do it himself', in addition to the clearly narrow focus reading 'I did it for him, not with him'. However, it is not central to my

argument whether accent on *for* conveys broad focus; what matters is that the two prominence patterns are distinct in form and in focus interpretation.

In addition to the location of the pitch accent, there are a number of other phonetic cues to stress in the phrase *for him*. When *for* has the main accent, it is likely to have a full vowel, and its final /r/ is likely to be realised even in non-rhotic varieties of English, since *him* is likely to be reduced to *'im* or *'m*. When *him* has the main accent, on the other hand, its initial /h/ is almost certain to be realised, and its vowel will be unreduced, whereas *for* will have a reduced centralised vowel and no /r/ in non-rhotic varieties. We might write these segmental variations informally as *FOR 'im* and *f'r HIM*.

The key question for the accent-first account is whether these segmental differences are merely concomitants of pitch accent, or whether they are cues to a phenomenon of 'stress' which is independent of pitch accent (see section 2.2.1). It is not hard to show that they can occur independently, and can be used on their own to signal relative prominence: that is, the phrase can be pronounced in a context where neither word is accented, and yet it is still possible to distinguish the relative prominence of the preposition and the pronoun, and the intended focus reading, by the segmental and durational cues alone. For example, we might have the following dialogues, in which the last pitch accent is on *program*, and the prepositional phrase is postnuclear and unaccented:

(13) (a) A: Bill says you haven't helped on his project very much.
B: I don't know what he's complaining about. I wrote an entire PROGRAM for 'im.

(b) A: Bill seems to think you've been giving priority to other people in the department.
B: I don't know what he's complaining about. I wrote an entire PROGRAM f'r him.

Despite the absence of any pitch accent on the prepositional phrase, the meaningful difference in relative prominence within the phrase can still be conveyed by the segmental differences.

These cases are closely related to the issue of so-called Second Occurrence Focus (SOF), which has been an important issue in the literature on the semantics of focus (e.g. Partee 1999). The classic cases of SOF involve focus distinctions in postnuclear deaccented stretches, such as the following example from Beaver *et al.* (2007: 253):

(14) A: Everyone already knew that Mary only eats vegetables.
B: If even Paul knew that Mary only eats vegetables, then he should have suggested a different restaurant.

B's reply involves two instances of what Rooth (1985, 1992) calls 'association with focus', namely the association of *even* with the nuclear accented *Paul*, and the association of *only* with the postnuclear *vegetables*. The latter is known as 'second occurrence' focus because the deaccenting of *vegetables* is seen as the result of its being the second mention of that discourse entity. The issue in the semantics literature has been whether SOF involves some degree of phonetic prominence on the relevant postnuclear word (in which case association with focus depends on prosody and should be regarded as part of the semantic representation of the sentence), or whether the relevant postnuclear word is 'completely deaccented' (which would mean that association with focus is driven purely by pragmatic considerations and need not be represented in the semantics). This whole issue is very carefully explored on the basis of new experimental data by Beaver *et al.* 2007, who show clearly that there are phonetic cues to the relative prominence of the SOF word, but that these cues cannot be regarded as involving a pitch accent in any ordinary understanding of the term. I suggest that the subtle phonetic cues to postnuclear prominence in SOF are entirely comparable to the segmental effects in postnuclear *for him* just discussed.

In any event, cases like (13) and (14) are inexplicable if we assume that focus is signalled only by pitch accent. Indeed, more generally, cases like these make it difficult to maintain the IPO/Bolinger view of the relation between stress and accent (section 2.2.1), according to which, stress is nothing but an abstract lexical property of certain syllables, realised in actual utterances – if at all – by pitch accent. Instead, they seem to require a theory of sentence stress based on relative strength in a metrical structure: in both (13a) and (13b) the focus distinction depends on whether *for* or *him* is the DTE of the constituent *for him*. As we saw in (12), the DTE of the constituent *for him* may also bear a pitch accent, but that is not essential for conveying the focus distinction. The presence or absence of pitch accent is a separate question.

This point of view still acknowledges the overwhelming association of primary accent with focus: it states that focus is signalled in the first instance by the location of DTEs, not accents, yet it also assumes that the DTE of any intermediate phrase will, by definition, have a primary accent. At the same time, however, it allows for cases of focus-without-accent, precisely because it treats the location of DTEs, not accents, as basic. Primary accents are often associated with focused words because focused words are often the DTE of intermediate phrases. Ultimately, however, accents do not respond directly to focus, but arrange themselves according to the demands of the metrical structure.

7.2.2 *Evidence for a metrical view of sentence stress*

Given the difficulties with an 'accent-first' view of the relation between focus and sentence stress, it seems worth developing the metrical alternative based on relative strength. This section outlines a possible stress-first theory that is consistent with the general ideas of metrical phonology. I will use Liberman–Prince metrical trees to exemplify metrical structure, but I leave open the question of whether this is the most appropriate representation. The main point is that some sort of metrical structure is at the heart of sentence prosody, and that pitch accents are essentially only one manifestation of the structure.

The first point to make is that, in many respects, the metrical interpretation of sentence stress makes the same predictions as an accent-first view. As we saw in section 2.2.2, the DTE of a short utterance is always accompanied by a pitch accent in any well-formed association of tune and text. This means that, given the metrical structures for *five FRANCS* and *FIVE francs* shown in (11) above, there must be a pitch accent on *francs* in one and on *five* in the other when the phrases are uttered in isolation. In short phrases, that is, the metrical stress-first view may seem descriptively equivalent to the accent-first view. However, even in short phrases the empirical claims of the two views differ subtly. The most obvious difference concerns the broad focus pattern *five FRANCS*. In the metrical view, it makes no difference whether there is an accent on *five* or not: the weak–strong relation between *five* and *francs* can hold in either case. That is, the metrical interpretation makes clear that the pattern of prominence that we have been writing as *five FRANCS* is not defined by the presence or absence of a pitch accent on *five*; what is important is only that *francs* is more prominent than *five*.

Not only can the metrical interpretation of sentence stress accommodate the presence or absence of secondary accent on *five* in the broad focus case, but in some sense it also predicts the existence of the ambiguity of the pattern *five FRANCS* (i.e. the ambiguity between the broad focus reading and the reading with narrow focus on *francs*). This ambiguity is an inevitable consequence of the limited range of structurally distinct metrical representations. That is, the fact that there are only two possible metrical relations in a two-word phrase, namely the two shown in (11), predicts that there are only two *phonologically* distinct prominence patterns on a two-word phrase. Since there are three interpretations – two narrow focus readings and a broad focus reading – then one of the two patterns *must* be ambiguous (although, as discussed in section 7.1.1 above, the patterns may be phonetically distinguishable through the use of paralinguistic emphasis). In an accent-first view, this ambiguity is essentially an accident; in a metrical account, it follows naturally from the representation.

Further justification for a metrical account of sentence stress comes from a consideration of deaccenting. In Ladd 1980a, I discussed several aspects of deaccenting in English that are difficult to explain solely in terms of the distribution of pitch accents and seem to demand a syntagmatic account. I argued explicitly that deaccenting is fundamentally a matter of metrical structure, and only secondarily of pitch accent. Specifically, I showed that there are certain phenomena of deaccenting that are puzzling under an accent-shift analysis, but which can be readily explained if we treat deaccenting as a *reversal of relative strength in a metrical tree*. The clearest example is the case of rightward shift of accent in deaccenting.

In classic cases of deaccenting, accent shifts to the left compared to its neutral or non-deaccented position:

(15) (a) A: Why didn't you read that article I gave you?
B: I can't read GERMAN.

(b) A: The only article on this is in German.
B: I can't READ German.

In some cases, however, deaccenting causes accent to shift to the right:

(16) (a) A: Where did you go just now?
B: I took the GARBAGE out.

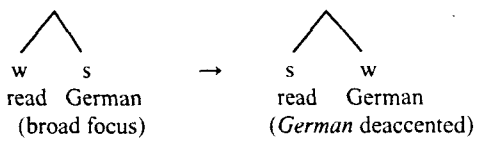
(b) A: What happened to all the garbage?
B: I took the garbage OUT.

(17) (a) A: Anything happen while I was out?
B: My PARENTS called.

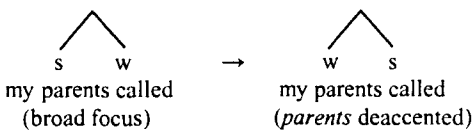
(b) A: Maybe we should call your parents and tell them.
B: My parents CALLED – they already know.

In Ladd 1980a, I suggested that both kinds of cases could be given a unified description in terms of reversed metrical nodes. So:

(18)



(19)



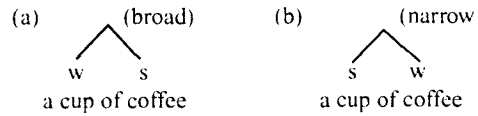
When one word is deaccented for pragmatic reasons, another word *must* be accented, because of the inherently relational or syntagmatic nature of prominence. But it is a matter of structure whether the word receiving the accent is to the left or the right of the deaccented word. It is difficult to account for this in terms of the linear shifting of pitch accents.

The presence of the discontinuous constituent *took . . . out* makes this analysis difficult to apply to (16) unless we resort to three-dimensional metrical trees – a topic that is unquestionably beyond the scope of this book to explore.[1] However, ignoring the discontinuity, the basic explanation advanced in the text for (15) and (17) works equally well for (16). In the neutral version, the argument (i.e. object noun) *garbage* is strong, and the predicate (i.e. discontinuous verb) *took . . . out* is weak. In the deaccented version, this prominence relation is reversed, so that *took . . . out* is strong. Within that strong constituent, in turn, *took* is weak and *out* is strong. In short, the primacy of arguments over predicates (see section 6.2.3) and the idea that deaccenting involves prominence reversal both apply just as well to this case as to (15) and (17). The only problem is diagramming the way in which these principles apply.

Finally, in addition to dealing easily with a wide range of deaccenting facts, a metrical interpretation of sentence stress also provides a natural account of the special status of the last accent discussed in section 7.1.2 above. Specifically, the special status of the rightmost accent in phrases like *five francs and a cup of coffee* follows naturally from the claim that broad focus is normally signalled by the relation w–s (weak–strong) rather than s–w (strong–weak), and from the fact that in metrical phonology we can define prominence relations *at any level of structure*. We start off with the w–s relation in *a cup of coffee* in order to signal broad focus in that constituent. As the focus constituent gets bigger and bigger, broad focus continues to be signalled by w–s relations at progressively higher levels of structure. This automatically makes the last accent of the largest phrase the strongest terminal element or DTE.

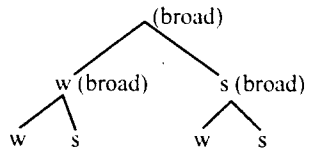
This analysis is illustrated in the following set of trees. First we show the broad focus and narrow focus pattern for the short phrase *a cup of coffee*:

(20)



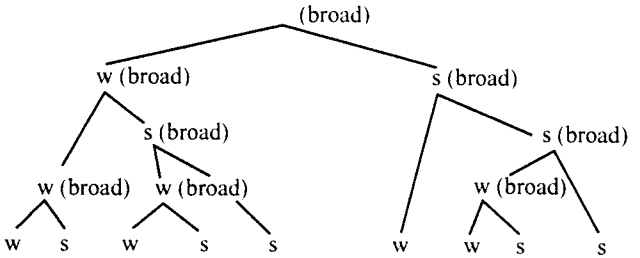
As we build progressively more complex phrases, the w–s relation between constituents continues to signal broad focus. Thus:

(21)



five francs and a cup of coffee

(22)



five francs seventy-five centimes and a cup of pretty tasteless coffee

In order to be interpreted as signalling broad focus, the strength relations in each constituent must be w–s.[2] As we saw in section 7.1.2 above, the resulting patterns can be realised in more and more different ways as the structure gets more complex, but the common element of all broad focus realisations is that *coffee* is the DTE and will always bear the last pitch accent.

More generally, the metrical analysis in all these examples expresses the implication that every phrase, and perhaps every utterance, has a single peak of prominence. This is the Trubetzkoyan idea that stress is a 'culminative' feature. As explained in more detail in the illuminating discussion by Beckman (1986: 19–27), a culminative feature is a feature that distinguishes one unit from the other similar units within some stretch of speech – for example, a feature that singles out one syllable in a word or phrase as most prominent. More or less by definition, no such notion of culminativeness figures in 'accent-first' generative accounts of sentence stress, nor generally in the FTA view. By

2 Some provision will of course have to be made in this statement for the fact that the constituent *seventy-five* may be affected in this context by rhythmic stress shift or 'iambic reversal' (Liberman and Prince 1977; Shattuck-Hufnagel, Ostendorf, and Ross 1994; Shattuck-Hufnagel 1995). A strong–weak pattern on this constituent in this context would not normally signal narrow focus on *seventy*, but would merely fit into a higher-level rhythmic regularity.

contrast, culminativeness can be readily accommodated in a metrical view, and indeed, one might argue, with Beckman, that the property of culminativeness is precisely what is formalised in the idea of the DTE and 'primary accent'.

7.2.3 *Sentence stress and prosodic phrasing*

In section 7.1, I argued that sentence stress is not essentially a matter of the location of pitch accents, but of *relative prominence within a metrical structure*, and that focus is signalled not directly by pitch accent, but by the location of locally defined primary stresses or DTEs. While I think the case against an accent-first view of sentence stress is more or less unanswerable, there is at least one aspect of primary stress that is somewhat problematic for a purely relational account of prominence. This problem – and the basis of a metrical solution – are the topic of this section.

The problem is seen most clearly in cases of 'double focus', which I mentioned at the beginning of chapter 6 and have since carefully avoided. Example (3) in chapter 6 is repeated here for convenience:

(23) I didn't give him SEVen EUros, I gave him FIVE FRANCS.

In both clauses, the noun phrases denoting sums of money each have two accents, indicating the presence of two points of focus or contrast in the sentence. The speaker in (23) is not simply opposing *five francs* to *seven euros* (as in a broad focus case like *I didn't give him the car keys, I gave him five francs*), but is explicitly opposing both *francs* to *euros* and *five* to *seven*.

Given the foregoing discussion of the nature of sentence stress, the additional focus on *five* and *seven* is difficult to account for. If the last accent is indeed more prominent in a metrical structure of the sort just discussed in section 7.2.2, then both the broad focus reading (with a secondary accent on *five*) and this double-focus reading will have the same weak–strong prominence relation:

(24)

five francs

If the two versions have the same metrical structure, then the theory has no explanation of why they should have different interpretations. Intuitively it seems clear that, in the terms suggested in section 7.1.1 above, the basis of the extra focus is that *five* has a primary accent rather than a secondary accent.

Somehow, the prominence is signalled by a difference of accent *type* – a paradigmatic distinction, rather than the purely syntagmatic contrasts to which we have attributed sentence prominence so far.

In order to accommodate this kind of case in a metrical theory of sentence stress, we will have to say that both *five* and *francs* bear primary accents, which in turn means that in some sense the two words are in *separate phrases*. The separate phrases can still be combined in a weak–strong relation. In ToBI terms there would be an intermediate phrase (*ip*) boundary (break index 3 and phrase accent) between the two words: each word is the DTE within its own *ip*, and each *ip* has its own intonational tune with a nucleus and a phrase accent. In an ordinary declarative reading of the phrase, there would be an L phrase accent in between the H* nuclear accents on the two words. This analysis accounts for the more prominent dip in pitch between the peaks that signals the presence of the double focus.

The reason this analysis is problematic, of course, is that it merely shifts the locus of the problem: instead of trying to define 'primary accent', we must now try to define 'intermediate phrase', so that we can state in a non-circular way when *five* is to be identified as a separate phrase from *francs*, and when the two words form a single phrase. However, there is reason to believe that this can be done. For example, Gussenhoven (1983b) tested listeners' ability to distinguish contextually appropriate prosodic patterns in pairs of sentences like the following:

(25) (a) He teaches linguistics. (subject–verb–object)
(b) He teaches in Ghana. (subject–verb–adverbial)

Gussenhoven's sentence-accent rules predict that when the sentences are expressing contextually new information, (25a) will constitute one 'focus domain' (roughly equivalent to intermediate phrase in the terminology used here), while (25b) will constitute two; when *teaches* is contextually given, there will be only a single focus domain in either case. Listeners' responses to these sentences in different contexts showed that the contextual appropriateness of the prosodic pattern was much easier to detect in cases like (25b) than in (25a). The explanation appears to be that in (25b), deaccenting *teaches* involves a change from two *ip*s to one, whereas in (25a) the sentence is a single *ip*, regardless of whether *teaches* is contextually given.

Further evidence for the validity of the general approach – trying to define primary accents by defining intermediate phrases – comes from considering an ostensibly unrelated problem. In English, as we saw in section 6.2.3, many

intransitive sentences have the main accent on the subject rather than on the verb:

(26) (a) The COFFEE machine broke.
(b) JOHNSON died.

As we also noted, this sentence stress pattern is less likely when the subject is very long, or when there are adverbial expressions intervening between subject and verb:

(27) (a) (?) The coffee machine in the ANTHROPOLOGY office broke down this morning.
(b) (?) Former President JOHNSON unexpectedly died today.

Instead, in a broad focus case we would be more likely to find accents on both subject and predicate, with the accent on the predicate generally perceived as the nucleus:

(28) (a) The coffee machine in the ANTHROPOLOGY office broke DOWN this morning.
(b) Former President JOHNSON unexpectedly DIED today.

Similarly, in section 6.2.1.3 we noted a comparable effect of sentence length in Romanian or Hungarian WH-questions, which would normally have the main accent on the WH-word at the beginning of the sentence, but which may have a later nucleus when the sentence is long.

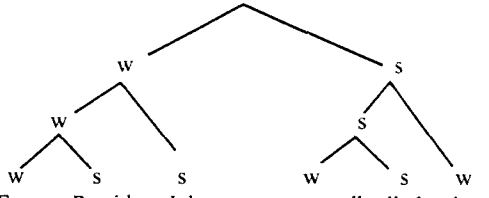
If we assume, with the 'normal stress' tradition, that there is a single nucleus for any sentence, no matter how long, and if we assume that there is a consistent syntactic-semantic basis for the location of the single nucleus, these shifts are somewhat puzzling. However, the puzzle disappears if we acknowledge that there are two accents in the longer cases, and that the accent on the predicate is perceived as strongest (i.e. as the nucleus) solely by virtue of its position, as we have already discussed extensively in sections 6.3.2 and 7.1.2. The effect of length comes about as follows. When the sentences are short, the predicate need not constitute an intermediate phrase of its own. Instead, subject and predicate are combined in a single intermediate phrase – a single focus domain, in Gussenhoven's terms – with a single primary accent. In English, under these circumstances, subject and predicate are in the relation strong–weak (see section 6.2.3), and the result is that the DTE of the single intermediate phrase occurs on the subject. By contrast, when the sentences are longer, it is difficult to treat the whole sentence as a single intermediate phrase, and the subject phrase and the predicate phrase therefore form separate intermediate phrases.

Within these phrases the expected strength relations apply, yielding DTEs (and accents) on *Johnson* and *died*:

(29)

At the highest level of structure, however, these two primary accents are in the relation weak–strong, exactly as in the case of *five francs and a cup of coffee* discussed above. This means that the DTE of the sentence – and hence the last pitch accent – is the DTE of the *predicate* phrase.

(30)



What these cases show us is that the overall prominence pattern of the sentence depends on its division into intermediate phrases. The intermediate phrases enter into a larger structure that defines their prominence relative to other intermediate phrases, but each intermediate phrase has a primary or nuclear accent. If we have two primary accents on syntactically closely related adjacent words, as in the double focus rendition of *five francs*, it is plausible to describe this by saying that each word forms its own intermediate phrase. In other words, double focus phrases do not oblige us to weaken our metrical theory of the relation between focus and sentence stress by adding a non-metrical distinction between primary and secondary accents; rather, they provide further evidence for the view that focus is signalled not directly by the location of specific pitch accents, but by the overall prosodic organisation of the sentence.

This general approach provides a ready explanation for sentences that seem to allow more than one 'neutral' pattern. Consider Halliday's humorous report (1970: 38) of the man who saw a sign saying *Dogs must be carried* in the London Underground 'and was worried because he had no dog'. Talking only in terms of the location of the last accent, we may say that this sentence

has two possible sentence stress patterns, neither of which involves narrow focus:

(31) (a) Dogs must be CARRIED.
(b) DOGS must be carried.

The intended meaning of the sign is something like 'If dogs are brought into the Underground, they must be carried', and the prominence pattern that conveys this meaning is (31a). The meaning that the worried man inferred from the sign is something like 'If you are in the Underground, you must carry a dog', and the prominence pattern that conveys this meaning is (31b). (This is the pattern that is appropriate for the sign *Shoes must be worn* commonly seen on the door of American shops and snack bars.) Gussenhoven (1983a) extensively discusses the difference of meaning that arises from the choice of one prominence pattern or the other: main accent on the verb gives a 'contingency' reading, and main accent on the noun gives an 'eventive' reading.

If we approach these from the point of view of accent – and particularly if we approach them from the point of view of the location of a single main accent – we will ask 'Why is the accent on *dogs* in one case and on *carried* in the other?' However, if we assume that accent reflects phrasing or prosodic structure, we will analyse the accentual difference as a difference between a version in which the whole sentence is grouped into a single prosodic phrase and one in which the sentence is divided into two phrases, *dogs* and *must be carried*. In the first case, where subject and predicate form a single intermediate phrase, there is a single primary accent; this accent is on *dogs*, in accordance with the principle of English that treats arguments as more accentable than predicates (section 6.2.3). In the second case, where subject and predicate each form a separate intermediate phrase, each has its own DTE and therefore its own accent (on *dogs* and *carried*, respectively). The accent on *carried* then becomes the main accent of the sentence for the reasons just sketched above.

This approach has the advantage of taking full account of the phonetic facts – in particular, the additional accent on *dogs* in (31a) – while at the same time allowing for the intuition that the accent on *carried* is the main accent of the sentence. In a sense it still leaves one central question unanswered, namely the question of why there should be two patterns, and why they should match up with interpretations as they do. Yet even that question becomes more tractable when the two patterns are seen as 'one domain versus two' rather than 'accent on subject versus accent on predicate'. The various discussions of these cases by Schmerling (1976), Gussenhoven (1983a), and Faber (1987) all suggest

that in the one-phrase pattern, the subject and predicate in some sense form a single unit of new information (Schmerling's 'news sentence', Gussenhoven's 'eventive reading'), while in the two-phrase pattern, the subject is in some sense separated out and presented as a reference point in the discourse and the really new information is in the predicate (Schmerling's 'topic–comment sentence', Gussenhoven's 'definitional reading'). Faber's terms 'integrative' and 'non-integrative' for the two patterns make the same basic point in a slightly different way.

Yet another advantage to looking at the expression of focus as a matter of metrical structure, rather than directly a matter of accent location, is that it gives us a better way to discuss universal features of the expression of focus. The extensive comparative discussion of focus and sentence stress in chapter 6 is obviously rather Eurocentric, because in many languages (particularly, but not exclusively, tone languages) it is difficult to identify a rigorously definable notion of 'pitch accent', even though there do seem to be prosodic cues to focus. If, on the other hand, we see prosodic phrasing as the ultimate basis of sentence stress, we may see that the correct way to pose questions about universals of the prosody–focus link is not 'Why is the main accent in this sentence on word X rather than on word Y?' but rather 'Why is this sentence divided up into prosodic phrases the way it is?'

Evidence for this way of looking at things comes from a comparison of European sentence stress data with observations on the location of phrase boundaries in a number of unrelated languages. For example, in their work on Japanese and Korean, Beckman and her students (e.g. Jun 1993; Venditti, Jun, and Beckman 1996) have proposed a notion of 'dephrasing' that produces effects strikingly similar to those brought about by deaccenting in European languages. This is not a peculiarity of Japanese and Korean; comparable influences of phrasing on the expression of focus in the Bantu language Chichewa were discussed by Kanerva (1990) and have recently been investigated in more detail for Chichewa and other Bantu Languages by Downing 2008. We can illustrate this general phenomenon with Korean examples from Jun 1993:[3]

(32) A: [saɾun-ənni] [irimi] [mwəni]
'What is cousin's name?' (lit. cousin name what)

B: [saɾun-ənni irimi] [suni-dʒi]
'Cousin's name is Suni.' (lit. cousin name Suni)

3 These examples are based on Jun's examples 7 and 8, pp. 197ff. I have simplified the glosses, and indicated phonological phrase boundaries by square brackets.

(33) (a) [na] [pap mək-illejo]
'I want to eat rice.' (lit. I rice eat-want)

(b) [na] [pap] [pəri-lejo]
'I want to throw out rice.' (lit. I rice throw-out-want)

(c) [na] [tol mək-illejo]
'I want to eat stone.' (lit. I stone eat-want)

In the question–answer sequence in (32), we see a straightforward analogue to contextual deaccenting: *irimi* ('name') forms a separate phrase when it is the point of A's question, but is 'dephrased' in B's answer. Example (33) is more complicated because it illustrates the interaction of structural and pragmatic influences. In (33a), 'rice' and 'eat' form a fairly predictable combination, and we observe the normal tendency for object–verb sequences to be realised as a single prosodic phrase; in English, a corresponding sentence could similarly phrase *eat* and *rice* together and completely deaccent *eat*. In (33b), though, when the rice is to be thrown out rather than eaten, verb and object form separate phrases in Korean; in English, to convey the same effect, the verb would tend to have its own accent. On the basis of these two examples alone, we might be tempted to equate accenting a word in English with treating a word as a separate phrase in Korean, so that these data could be reconciled with a basically Bolingerian highlighting approach to such matters. But (33c) shows that things are not that simple: though the combination of 'eat' and 'stone' could scarcely be called predictable, the tendency is nevertheless to group object and verb in a single prosodic phrase. Given the essentially structural fact that object–verb sequences are normally realised as a single phrase, a two-phrase realisation would lend undue prominence to 'eat', and would therefore inappropriately de-emphasise 'stone'. There is thus a kind of trade-off between structural tendencies and pragmatic expectations, rather like what we saw in the cases from English and Icelandic discussed in section 6.2.2.4.

Furthermore, there may be a relation between the extent to which a language tolerates deaccenting and the way in which the language normally divides utterance into prosodic units. Recent work suggests that there are differences between languages in their tolerance of long prosodic phrases: in some languages (e.g. Spanish, Elordieta *et al.* 2003; Egyptian Arabic, Hellmuth 2007), almost every content word seems to form its own intermediate phrase, whereas in others (e.g. Portuguese, Elordieta *et al.* 2003), two or more content words can readily be combined into a single phrase. There are obviously difficulties in defining exactly what we mean by intermediate phrase, but the use of comparative data by Elordieta *et al.* – as with the comparative sentence stress data in

chapter 6 – makes it harder to argue that the differences are not genuine. It is as if Portuguese (like English) tolerates a fairly deep prosodic structure, with one content word subordinated to another, whereas Spanish and Egyptian Arabic prefer a fairly flat prosodic structure consisting of a string of short intermediate phrases. It may or may not be coincidence that Portuguese is also more accepting of the West Germanic type of contextual deaccenting (section 6.2.2.1) than either Spanish or Egyptian Arabic. However, it is plausible to see a connection between the two typological facts, because – on the interpretation proposed here – deaccenting is phonologically a matter of prosodic subordination. This is a topic for future typological research.

In any case, it would obviously be desirable for a universally valid theory of the link between prosody and focus to be able to relate observations about phrasing to observations about sentence stress. If we treat the relation between focus and sentence stress in English as indirect, we are in a better position to draw such universal connections. That is, let us assume that in both English or Dutch and Korean or Chichewa, focus is systematically signalled by prosodic structure. In Korean or Chichewa, the main observable consequence of differences of prosodic structure is the presence or absence of accentual phrase boundaries (hence 'dephrasing'). In English or Dutch, the most salient consequence of such differences is the presence or absence of pitch accents (hence 'deaccenting'). But dephrasing and deaccenting, in the view proposed here, are just different surface symptoms of the same deeper structural effects.

8 *Prosodic structure*

The treatment of the link between sentence stress and focus in the previous chapter sets the stage for this final chapter, in which I explore how the 'metrical' and 'autosegmental' aspects of intonation fit together. Once we adopt the idea that a culminative metrical structure is central to understanding the relation between sentence stress and focus, we find that the same idea sheds light on other problems in intonational phonology. The goal of this chapter is to show, for several different and superficially unrelated issues, that explicit recognition of metrical structure in intonational phonology helps make sense of several long-standing puzzles, and clears away some conceptual problems that have held back the field for too long.

8.1 The structure of intonational tunes

8.1.1 Prenuclear accents in tune–text association

It is widely assumed that a language's intonational phenomena can be classified into contour *types* or 'tunes'. Many descriptions of many European languages contain references to 'neutral declarative' intonation, 'interrogative' intonation, and the like. Some descriptions of English go considerably further than this, positing specific tune types, like the 'contradiction contour' (Liberman and Sag 1974) or the 'surprise–redundancy contour' (Sag and Liberman 1975). In the same way, the IPO description of Dutch identifies such tunes as the 'hat pattern' and the '3C', while Delattre's classification of French tune types (Delattre 1966) includes such tune types as 'major continuation', 'minor continuation', and 'implication'. Regardless of how many such types are recognised, one central goal of theories of intonational phonology is to be able to provide an explicit phonetic characterisation of all the tunes of a given language. In particular, we want to be able to make explicit predictions of how a given tune will be realised when it is applied to different texts.

A crucial ingredient of any such theory is an account of *tune–text association*. As we saw in section 2.2.2, this is the term employed by Liberman (1975) to

describe the principles according to which the abstract phonological elements of the contour are lined up with the segmental features of the utterance. I do not of course mean to suggest that the principles that Liberman identified as principles of tune–text association were absent from earlier descriptions, but only that such principles operated almost entirely out of the awareness of researchers. The only important exception to this generalisation known to me is 't Hart and Collier's discussion of contour identity (1975), which develops very similar ideas to Liberman's.

Consider the application of 'the same' contour to utterances with different numbers of syllables. Suppose, for example, we identify for English an 'emphatic' statement contour with an ordinary H* accent near the beginning of the utterance, and a raised or emphatic H*_L_L% or L+H*_L_L% sequence (here annotated ad hoc with an up-arrow ↑ before the H*) at the end:

(1) (a) H* L+↑H* LL%
Do it now.
(b) H* L+↑H* LL%
Her mother's a lawyer.

Intuitively we are dealing with the same contour in both cases, despite the fact that the two utterances have different numbers of syllables. Pierrehumbert's tonal analysis (or indeed any analysis that treats the intonation contour as a sequence of distinctive tonal events at particular points in the utterance) gives expression to this intuition: in both cases the contour is represented abstractly as H* ... L+↑H*_L_L%. The association of the H* accents with *do* and *now* and the stressed syllables of *mother* and *lawyer* follows in a lawful way from their relative prominence. The autosegmental representation of the tonal events and the metrical account of how the tonal events are associated with syllables give us a clear and explicit basis for describing the two contours as 'the same'.

However, there are many cases where an account of tune–text association based on current AM intonational phonology falls short of capturing our intuitive understanding of what it means for two utterances to have the same intonation. One of the most serious problems involves prenuclear secondary accents. As we have just seen, it is straightforward to represent a contrastive statement contour H* ... L+↑H*_L_L% which can apply to texts with different numbers of *syllables*, but a problem arises whenever contours contain different numbers of *accents*. In this case, standard AM assumptions provide us with no phonological abstraction that will allow us to refer to the resulting contours as 'the same', even though intuitively that seems appropriate. This can be seen from the following sentences.

(2) (a) (two prenuclear accents)
H* H* L+↑H* LL%
Mary McKelvie's a lawyer.
(b) (no prenuclear accent)
L+↑H* L L%
Now!

Intuitively, these contours are the same as those in (1). The surface variations in (2a) and (2b) are just as lawful as those between (1a) and (1b): they are determined by the number and type of syllables in the text. Ultimately, then, they should be covered by a general theory of tune–text association; but given current AM assumptions, the representations H* ... H* ... L+↑H*_L_L% (2a) and L+↑H*_L_L% (2b) must count as different, both from each other and from the H* ... L+↑H*_L_L% in (1a) and (1b).

The same problem can be seen more clearly in the following pairs of sentences (from Ladd 1986):

(3) (intonation nuance: relatively neutral)
H* H* LL%
(a) I read it to Julia.
H* H* H* LL%
(b) I wanted to read it to Julia.

(4) (intonational nuance: '... and that's that')
H* !H* LL%
(a) I read it to Julia.
H* !H* !H*LL%
(b) I wanted to read it to Julia.

(5) (intonational nuance: '... as you ought to know')
L* H* LL%
(a) I read it to Julia.
L* ↓L* H* LL%
(b) I wanted to read it to Julia.

Both members of each pair appear to have the same intonation, in the sense that both convey the same pragmatic force suggested by the informal glosses of the intonational nuances. The number of accents preceding the final H*_L_L% seems to depend solely on the number of accentable syllables; and the categorical identity of those accents (H*, H*, followed by downstep, L*) seems to represent a *single linguistic choice*, regardless of whether there are one or

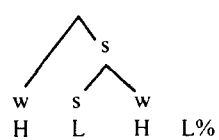
two such accents. Somehow, we want to be able to treat the final H*_L_L% sequence, which occurs only once, separately from the accent or accents that precede it, and to treat the exact number of preceding accents as a detail that is subject to predictable variation. That is, we want to be able to treat the phonological specification of the 'surprise–redundancy contour' in (5) as something like 'optionally one or more L* accents followed by an H*_L_L% sequence'.

8.1.2 *Metrical structure for tunes*

The problem of how to abstract away from the number of prenuclear accents is comparable to the problem of postnuclear accents discussed in section 4.1.4. As we saw, Grice, Ladd, and Arvaniti 2000 proposed that Pierrehumbert's notion of 'phrase accent' should be defined as a peripheral tone that can be associated either with an intermediate phrase boundary or with a postnuclear stressed syllable if one is present. But this causes a number of theoretical difficulties given current assumptions: even if we accept the notion of 'postnuclear accent', there remains the fact that there is no obvious basis in AM theory for a tonal element that can surface as a pitch accent or an edge tone depending on the context. In the same way, there is no provision in the theory for a prenuclear tonal element that, depending on the context, can surface once, twice, or not at all.

The essential structural feature of both prenuclear and postnuclear accents is that they are secondary relative to the primary nuclear accent. The structural feature that distinguishes prenuclear from postnuclear accents is that the latter are *obligatory*: prenuclear accents can be completely absent, but the postnuclear 'phrase accent' is always manifested phonetically, either as a postnuclear accent or a boundary tone. This suggests that the difference between nuclear, postnuclear, and prenuclear tones can be treated as a matter of the relation between them in a metrical or other syntagmatic structure. For example, we might indicate the tonal structure of the 'Eastern European Question Tune' discussed by Grice, Ladd, and Arvaniti (2000) (cf. sections 2.5 and 4.1.4) as follows:

(6)

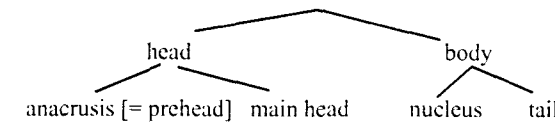


This shows the obligatory low nuclear accent as strong relative to the following postnuclear peak, and treats the combination of the nuclear and postnuclear

tones as a constituent which is strong relative to the prenuclear high accent; it treats the boundary tone as being in some way outside the main structure of the tune.

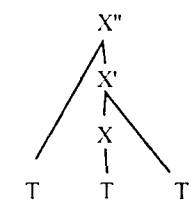
Just such a constituent structure for tunes was proposed by Liberman (1975: sect. 2.4.3) and developed by Pierrehumbert and Beckman 1988. The similarity of this idea to Grice's proposals for the internal structure of pitch accents discussed in section 4.1.3 will also be obvious. In fact, the idea that tunes have a hierarchical constituent structure of some sort goes back many decades. For example, one early version (Chao 1932) of the traditional British head–nucleus–tail analysis (discussed in section 4.1.1) suggested a hierarchical structure that could be represented in tree form as follows:

(7)



Updating the traditional British analysis in terms of the X-bar theory of constituent structure (e.g. Kornai and Pullum 1990), we might propose the following general structure for intonational tunes:

(8)



In this structure, the nuclear accent is analysed as the 'head' of the tune.[1] The British tradition's tail – redefined in AM terms as the phrase accent that surfaces variously as edge tone or as postnuclear accent – is analysed as the 'complement'. The prenuclear accent (the main accent of the 'head' of the British tradition) is treated as a 'specifier' or as an 'adjunct'. The final boundary tone is seen (as it was by Liberman 1975) as being outside the main structure of the tune.

1 The terminological confusion between 'head' in the X-bar sense and 'head' in the traditional British sense of 'prenuclear stretch of contour' is unfortunate but cannot be helped; the only solution appears to be to avoid the term 'head' in discussing intonation altogether.

Such a structure is consistent with many of the observations that we have just been discussing:

(1) A tune has one element – the nucleus – which is in some sense its central or most prominent point.
(2) A tune has a constituent structure in which the most major break is that between the nucleus and all that precedes; the distinctive postnuclear elements of a tune are more closely bound to the nucleus than are the prenuclear elements.
(3) Postnuclear elements in at least some languages may surface either as accents or as edge tones, depending on the metrical structure of the segmental material to which they are associated.
(4) The prenuclear element in a tune is – or may be – a single linguistic choice. The occurrence of multiple prenuclear accents depends on the metrical structure of the segmental material to which the tune is associated.

By itself, of course, the structure proposed in (8) does not actually describe in detail how prenuclear and postnuclear elements are associated with segmental strings of different metrical structure. (Indeed, even the association of the nuclear tone with the most prominent syllable in the metrical structure of the sentence is not clearly shown.) Obviously, therefore, in order to make any such proposal 'work', formal mechanisms will have to be provided to describe these effects. For example, there might be tonal specifications of some sort on the parent intonational phrase node, specifications that percolate down to accent-bearing daughter nodes, yielding slightly different tonal strings depending on the metrical structure of the intonational phrase. But whatever mechanism is proposed, the idea that tunes have hierarchical structure has clear implications for a general theory of tune–text association. In particular, the conclusion seems inescapable that the tune is a property of a phrase, and is thus *more abstract than a string of tones*. The elements of tunes are abstract tones, and they are not intrinsically either accent tones or edge tones (i.e. not intrinsically either starred or unstarred). Rather, what they are is intrinsically nuclear, prenuclear, or postnuclear. The prenuclear tones may surface as one or more accents, or may be deleted or truncated. The postnuclear tones often surface as edge tones, and form a close group with the nuclear tone, but they may also surface as accents under certain circumstances.

Two consequences of the notion of an abstract tune are worth mentioning here. The first is that it makes it possible to describe the restrictions on the pitch

patterns of various kinds of tags. As has often been noted, for example, the pitch contour on tags such as *she said* or *he replied* is in some sense determined by (or at least consistent with) the pitch contour on the main part of the sentence:

(9) H* H H H%
(a) 'Are you coming?' she asked.
L+H* L L L%
(b) 'Get out of here!' he yelled.
L*+H L LH%
(c) 'I don't t h i n k so,' he said.

If the main sentence has a high-rising nucleus, the intonation on the tag continues high; if the main sentence has a falling nucleus, the intonation on the tag continues low; and – importantly – if the main sentence 'should' have a fall–rise nucleus, one possible pattern is for the falling accent to show up on the main sentence and the rise at the end of the tag. (Cf. the similar behaviour of intonational tags in French, illustrated in examples (16–20) in chapter 3.) However, as has also often been noted, the tag generally has some sort of accent: there is a clear difference in the location of the valley in, for example, *he replied* and *he commented*, which have the greatest prominence on their final syllable and their antepenultimate syllable respectively. This kind of phenomenon can be readily accommodated in a descriptive framework in which the postnuclear tone is not intrinsically either a pitch accent or an edge tone, and is associated with the most prominent syllable available. Compare this analysis of intonational tags to the treatment proposed by Beckman and Pierrehumbert (1986: 293–8), in which the subordination of the tag is treated as a matter of pitch range compression (cf. section 5.2.5). Beckman and Pierrehumbert dismiss the intuition that the tag is deaccented as a kind of illusion induced by the difference in pitch range, and they leave unexplained the restrictions on which tones can occur on the tag.

A second consequence may be more far-reaching. If the tones in the abstract representation of a tune are not intrinsically starred or unstarred – if their status as accent tones or edge tones depends on the metrical structure of the text to which the tune is associated – it follows that sentence-level prominence must be defined in the metrical structure. The distribution of starred pitch accents is derived, not specified directly. The discussion in this section therefore converges with the other evidence for the same general conclusion presented in chapter 7.

8.2 Aspects of prosodic structure

8.2.1 *On defining prosodic boundaries*

It is universally assumed that one of the functions of prosody is to divide up the stream of speech into chunks or phrases of one sort or another – for the moment, informally, let us call these chunks prosodic phrases or ProsPs. Despite the apparent universality of the chunking function, however, ProsPs and ProsP boundaries are remarkably difficult to define and to identify consistently. ProsP boundaries seem to take on a bewildering variety of manifestations, from a clear pause accompanied by a local F_0 fall or rise, to a subtle local slowing or pitch change that defies unambiguous definition. As a result, there is often disagreement about whether a particular ProsP boundary is or is not present, and definitions of ProsP boundaries in the literature are frequently circular or vague. For more discussion of this problem, see Crystal (1969: sect. 5.4); Liberman (1975: 9ff.); and Ladd (1986: sect. 1.1).

In my view, the difficulty of identifying ProsP boundaries has less to do with the inherent subtlety of the phonetic cues than with the fact that their definitions involve conflicting criteria. As with the phonetic definition of stress (section 2.2), the problem is primarily theoretical. ProsPs are supposed to be set off by audible boundaries: if ProsP boundaries were not audible, then much of the point of the chunking function would be lost. At the same time, ProsPs are frequently assumed to have an internal prosodic structure of some sort – an intonational tune, a DTE, and so on. The details of the supposed internal structure vary from analysis to analysis, but the effect is the same: the assumption of internal prosodic structure creates a potential for *theoretically incompatible observations*. We may find stretches of speech that appear to be delimited by prosodic boundaries, but do not exhibit the expected internal structure; conversely, we may find stretches of speech that appear to have the internal structure of a ProsP, but whose edges are not marked by audible boundaries.

The most important – and most complex – conflict of criteria arises from the twin assumptions that (a) the division of sentences into ProsPs in some way reflects syntactic, semantic, or discourse constituency, but that (b) prosodic structure is somehow fundamentally simpler than syntactic structure. Exploring the relationship between syntactic/semantic and prosodic structure has been a major area of research in prosody for several decades (e.g. Halliday 1967b; Langendoen 1975; Nespor and Vogel 1982, 1983, 1986; Selkirk 1984, 1986, 1995; Ladd 1986; Chen 1987; Kubozono 1987; Steedman 1991, 2000; Lambrecht 1994; Croft 1995; Truckenbrodt 1999; Wagner 2005). Obviously, the assumed existence of a broadly grammatical system underlying prosodic phrasing places still another potentially conflicting constraint on where we may observe ProsP boundaries (and indeed, a major part of what the works just referred to attempt to do is to explain – or explain away – 'mismatches' between prosodic structure and syntactic/semantic constituency). If we hear an audible break in a syntactically or semantically 'impossible' location, we may be tempted to say that it is a hesitation rather than a ProsP boundary; conversely, if we fail to observe a clear boundary where our rules lead us to expect one, we may be tempted to conclude that one is present anyway, but that it is hard to hear.

This conflict of criteria is exacerbated by what seems to me to be fuzzy thinking on what it means to define ProsPs and prosodic domains more generally. I take it as self-evident that it is important for prosodic domain types to have phonetically explicit definitions. However, several people with whom I have discussed these issues seem to assume that the various prosodic domain types are *defined* by descriptions of how syntactic structure is mapped onto prosodic structure – such as those of Nespor and Vogel 1986 or Selkirk 1986 (e.g. a phonological phrase is defined as the head of a maximal projection and everything to its left). In my view, it makes no sense to treat accounts like Nespor and Vogel's or Selkirk's as definitions; rather, they are hypotheses, predictions about the correspondence between one type of independently definable structure and another. For example, we may want to define a phonological phrase as the domain whose DTE bears a nuclear pitch accent, or as the stretch between two phrase accents, or in some other way; but the adequacy of the definition must be evaluated in the first instance on phonological and phonetic grounds. It is then an empirical matter whether Nespor and Vogel's account or Selkirk's account successfully predicts the correspondence between syntactic maximal projections and phonological phrases as independently defined. Unless the syntactic and the phonological structures are defined in their own terms, the whole exercise becomes purely circular.

Considered purely as a problem of phonetics and phonology, the boundaries of ProsPs and other prosodic domains are actually not especially difficult to define and identify. In the ToBI transcription system, for example, the definitions of boundaries are expressed quite clearly in phonetic terms, and there is good inter-transcriber agreement on the location of accents and boundaries. (Admittedly, as noted in section 3.2.1, the system itself has an explicit mechanism – break index 2 – for dealing with theoretically incompatible observations.) In the study by de Pijper and Sanderman (1995) already cited (section 3.1.5), non-expert listeners showed good agreement in rating the 'strength' of

prosodic boundaries on a ten-point scale, even when the segmental content of the utterances was rendered unintelligible. In the face of results like these, it is hard to maintain that prosodic boundaries are especially elusive or mysterious. Rather, the problem arises from conflicting criteria when we attempt to reconcile clear phonetic and phonological definitions with independent ideas about prosodic structure and its relation to syntax. The main source of the difficulty is the so-called Strict Layer Hypothesis – the idea that prosodic structure, unlike syntactic structure, is of fixed depth.

8.2.2 *The indeterminate depth of prosodic structure*

The differences between prosody and syntax – and the problems these differences pose for phonetic definitions of prosodic constituents – are illustrated by the well-known case of the children's poem *The house that Jack built*, discussed in this connection by Chomsky and Halle (1968: 371ff.). As Chomsky and Halle pointed out, the syntactic structure of each verse is indefinitely right-branching:

(10) [This is [the dog that chased [the cat that killed [the rat that ate [the malt that lay in [the house that Jack built]]]]]]

The prosodic structure, however, involves a succession of phrases in which this depth does not appear to be reflected in any way.

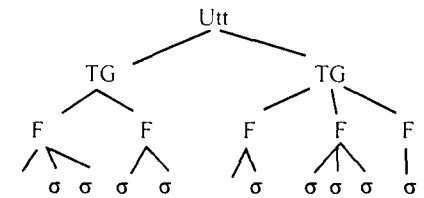
(11) | This is the dog | that chased the cat | that killed the rat | that ate the malt | that lay in the house that Jack built |

Moreover, as can be seen, the boundaries between these prosodic phrases 'come in the wrong place' from the point of view of the syntax: not at the beginning of each noun phrase, but somewhere in the middle. Such mismatches are generally taken as the inevitable result of attempting to map one kind of structure (indefinitely branching syntactic structure) onto another fundamentally simpler kind (the intuitively 'flatter' or 'shallower' prosodic structure); this point of view is reflected especially clearly in the work of Nespor and Vogel (1986), Selkirk (e.g. 1986), and Truckenbrodt (1999).

There is widespread agreement that the prosodic structure of utterances involves a hierarchy of well-defined domain types (syllables, phrases, etc.), such that the boundaries at each level of structure are also boundaries at the next lower level. This idea was first explicitly discussed by Halliday (1960, 1967a). For example, in Halliday's analysis of English phonological structure, utterances are composed of tone groups, tone groups of feet, feet of syllables, and syllables of phonemes. This means that every tone group boundary is

simultaneously a foot boundary and a syllable boundary. Such a structure can be represented in a tree diagram:[2]

(12)



// ^ When I / got there // ^ she'd / already / left //

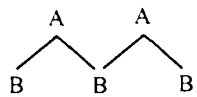
More recent investigators (notably Nespor and Vogel 1986 and Pierrehumbert and Beckman 1988) have proposed rather different inventories of categories or domain types, but the principle that they can be arranged in a hierarchy of size or inclusiveness is accepted by nearly everyone.

One property of trees like (12), much discussed by Halliday and his followers and critics (Huddleston 1965; Halliday 1966; Matthews 1966), and more recently by, for example, Selkirk (1984) and Nespor and Vogel (1986), is that they are *non-recursive*. No node may dominate another node of the same category. This property was incorporated into several theoretical discussions of the 1980s in the form of a specific constraint on prosodic structure that Selkirk (1984) dubbed the 'Strict Layer Hypothesis'. According to the Strict Layer Hypothesis (SLH), there is a hierarchy of prosodic domain types such that, in a prosodic tree, any domain at a given level of the hierarchy consists exclusively of domains at the next lower level of the hierarchy.

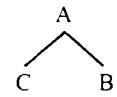
This abstract statement translates into some of the concrete prohibitions shown graphically in figure 8.1. More importantly for the point under discussion, it entails that *prosodic trees are of fixed depth*. Any tree constructed

2 In the text of this example, the tone group boundaries are marked by // and the foot boundaries by /, while ^ marks a 'silent ictus'. The latter is a hypothetical beat at the beginning of tone-group-initial feet; it stands in place of the stressed (Halliday's 'salient') syllable with which feet normally begin. In the case of the foot ^ *she'd*, the silent ictus is said to be manifested phonetically by a pause whose duration is governed by the rhythm of the sentence as a whole (see Halliday 1970); there is some experimental evidence for this idea (Scott 1982). However, in the case of the foot ^ *When I*, the motivation for the silent ictus is purely theoretical. Utterance-initial unstressed syllables must be assigned to a foot, and a foot by definition must have an 'ictus' or salient beat; since there is no observable ictus, it must be silent. Nowadays cases like this might be discussed in terms of 'extrametricality' (e.g. Hayes 1981; Pierrehumbert and Beckman 1988: ch. 6) or 'degenerate feet' (Halle and Vergnaud 1987), but the 'silent ictus' nevertheless remains a nice example of the kind of construct that is likely to arise from theoretically incompatible observations.

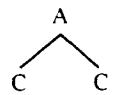
(a) No multiple domination



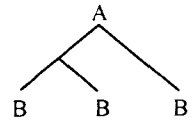
(b) No heterogeneous sisters



(c) No skipping of levels



(d) No unlabelled nodes



(e) No recursion

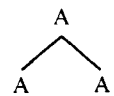


Figure 8.1. *Some consequences of the Strict Layer Hypothesis. The category labels A, B, and C in the diagrams are intended to represent prosodic domain types (intonational phrase, phonological word, foot, syllable, etc.), where A is a larger domain than B, and B is a larger domain than C.*

according to the constraints of the SLH as just stated can have only as many levels of structure as there are distinct domain types: for example, in Halliday's model shown in (12), with four domain types *Utt*, *TG*, *F*, and σ, every σ node will always be exactly three levels down from the root *Utt* node. The SLH thus gives expression to the intuitive 'flatness' of prosodic trees by equating flatness with fixed depth. By restricting the set of prosodic trees to those of fixed depth, it also (as just noted above) makes it inevitable that there will be certain types of syntactic structures (such as repeated right branching) that cannot be reproduced in the corresponding prosodic trees. It thereby provides a reason why there should be 'mismatches' between prosody and syntax.

Once it was explicitly formulated, the SLH was widely taken to be self-evident. Four years after Selkirk formulated it as a 'useful working hypothesis' (1984: 26), Pierrehumbert and Beckman (1988: ch. 6) incorporated it into their formal system as an axiom in the literal sense – and making no reference to Selkirk, Nespor and Vogel, or the term 'Strict Layer Hypothesis'. Both Selkirk and Nespor and Vogel revised earlier analyses that did not conform to the

SLH (e.g. Selkirk 1980; Nespor and Vogel 1982, 1983) and replaced them with ones that did (e.g. Selkirk 1984; Nespor and Vogel 1986). Moreover, in some important sense, the 'grid-only' models of prosodic structure proposed by Prince (1983) and developed by Selkirk (1984) may be said to differ from the grids of Liberman and Prince (1977) precisely by incorporating the SLH: Liberman and Prince's purely relational grid is of indeterminate depth, whereas the Prince–Selkirk grid has exactly five well-defined levels.

But the rapid acceptance of the SLH can hardly be attributed to overwhelming empirical evidence that prosodic structure is of fixed depth, or to clear data supporting the prohibitions in figure 8.1. On the contrary, the SLH clearly causes empirical problems. For example, both Hyman, Katamba, and Walusimbi (1987) and Chen (1987) point to bracketing paradoxes that arise from assuming the SLH in languages with complex tone sandhi. Some of my own work (esp. Ladd 1986, and the first edition of this book) has drawn attention to the difficulty of reconciling the SLH with phonetic evidence for degrees of boundary strength. As a result, many researchers now accept that some form of recursion and/or indeterminate depth of structure must be incorporated into our understanding of prosodic structure (cf. e.g. Dresher 1994; Frota 1998; Féry and Truckenbrodt 2005; Wagner 2005). Many details, though, remain to be worked out.

8.2.3 *Boundary strength*

It is worth discussing the issue of boundary strength in more detail. It has been shown in several studies that clear ProsP boundaries may have different acoustic properties that in some way reflect the structural strength or depth of the boundary (see e.g. Cooper and Sorensen 1981; Thorsen 1985, 1986; Ladd 1988). It has also been found, as just noted, that listeners are able to detect such differences reliably when segmental content is acoustically masked (de Pijper and Sanderman 1995). Indeed, before the SLH was formulated, the notion of boundary strength was taken for granted in some early work on the relation between syntactic and prosodic structure (e.g. Cooper and Paccia-Cooper 1980; Gee and Grosjean 1983). However, if boundary strength is to be defined on prosodic trees that conform to the SLH, there is a conflict between the SLH's prediction of fixed depth of structure and the phonetic evidence that boundary strength is a potentially open-ended scale.

The nature of the conflict can easily be demonstrated. In (11) I marked the last phrase of the utterance as *that lay in the house that Jack built*. What I really hear when I say the verse to myself is that there is another boundary

after *house*, but that it is weaker or more subtle than the ones I marked in (11). Under the usual assumptions about prosodic phrasing, there is no good way to indicate this percept of a 'weaker' boundary; either it is there (and I perceive it as weaker because I am aware of syntactic differences between it and the other boundaries) or it is not there (but I perceive it anyway because I know it is there syntactically).

There are only two ways to reconcile evidence for phonetic cues to boundary strength with the SLH. One is to interpret different boundary strengths as different *probabilities* with which a given syntactic boundary will be marked prosodically. This is the approach taken in Pierrehumbert and Liberman's reinterpretation of Cooper and Sorensen, discussed in section 1.3.1. The other is to express the perceived weakness of a given boundary directly, by saying that it is a *different kind of boundary* – specifically, a lower or more minor boundary. Under the SLH, as a moment's reflection will make clear, the only possible differences of boundary strength are differences of boundary type.

The idea that there are two different kinds of ProsPs has been proposed independently a number of times over the past half-century. These proposals include those of Trim (1959; major and minor tone groups), O'Connor and Arnold (1973; single and double bar boundaries), and Beckman and Pierrehumbert (1986; intonation phrases (*IP*s) and intermediate phrases (*ip*s)). In all cases, the theoretical proposal amounts to distinguishing between big ProsPs and little ProsPs, with big ones consisting of one or more little ones – as required by the SLH. Unfortunately, there is no clear correspondence among the various proposals for two levels of ProsP. For example, as I noted in Ladd 1986, Trim's big ProsPs are roughly comparable to O'Connor and Arnold's little ProsPs. If we were to credit the observations of both, we would therefore end up with three levels of ProsP; and detailed comparison with other proposals could lead us to posit even more. In order to avoid positing too many different types of boundaries, investigators more often resolve the implicit conflict of criteria by simply ignoring certain boundary cues, and falling back on the 'elusiveness' of prosodic phrasing as an explanation for their difficulties. This is, in effect, what I did when I marked the location of ProsP boundaries in example (11).

So long as observations of boundary strength differences are impressionistic, it is possible to dismiss them as syntactically induced illusions, or to explain them away in other ways, and maintain the SLH. But there is an increasing amount of instrumental evidence for such differences, which can be exemplified by the results of a study that I carried out some years ago (Ladd 1988), more recently replicated for a slightly different structure in German by Féry and

Truckenbrodt (2005). My study involved sentences of the form *A and B but C* and *A but B and C*, where *A*, *B*, and *C* are clauses of roughly similar syntactic and accentual structure, such as:

(13) (a) Warren is a stronger campaigner, and Ryan has more popular policies, but Allen has a lot more money.

(b) Warren is a stronger campaigner, but Ryan has more popular policies, and Allen has a lot more money.

The most natural interpretation of these sentences treats the *but* boundary as stronger – that is, in (13a) it opposes the conjunction of *A and B* to *C*, while in (13b) it opposes *A* to the conjunction of *B and C*. Graphically:

(14) (a)

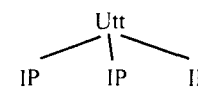
A and B but C

(b)

A but B and C

In multiple readings of these sentences by four speakers, I found evidence of F_0 declination during each clause and across the entire sequence of three clauses; each clause ended with a clear boundary tone. It therefore seems appropriate to treat each clause as an intonational phrase, and the whole thing as an utterance. Given standard assumptions about prosodic structure, including the SLH, we might therefore diagram the structure of these utterances as follows:

(15)

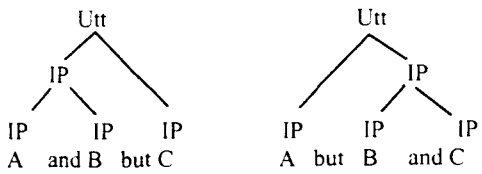


In line with the standard motivations for the SLH, this structure is 'flatter' than the structures in (14).

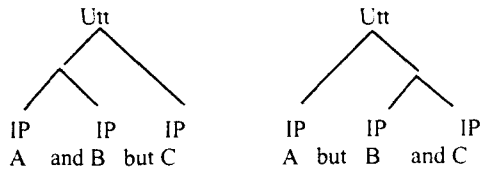
In addition to the clear phonetic evidence for dividing up the utterance into three intonational phrases, however, I also found phonetic differences that reflect the hierarchical organisation shown in (14). Specifically, the initial peaks of clauses *B* and *C* were higher after a *but* boundary than after an *and* boundary. Moreover, the pauses preceding *but* boundaries were by and large longer than those preceding *and* boundaries. That is, both the F_0 cues and the pause-duration cues agree in signalling that the *but* boundaries are 'stronger'.

Unfortunately, the most obvious ways of representing this difference of boundary strength are both ruled out by the SLH:

(16) Recursive node

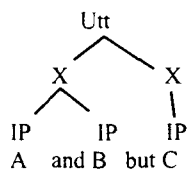


(17) Unlabelled node



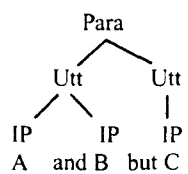
This is because, as we noted earlier, the only differences of boundary strength permitted by the SLH are differences of boundary type. In order to represent the hierarchical organisation of (14) without violating the SLH, we must come up with a new domain type for the intermediate layer of structure between the root *Utt* node and the string of three *IP* nodes:

(18)



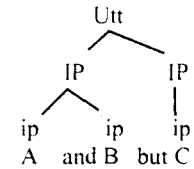
If we restrict our analysis to recognised domain types, we must, as it were, slide them up or down the tree in order to accommodate the demands of the SLH. For example, we might treat the conjunction of clauses *A* and *B* as an utterance, and analyse the whole structure as a paragraph:

(19)



Alternatively, we could treat the conjunction of clauses *A* and *B* as an *IP*, which then forces us to treat the individual clauses as intermediate phrases (*ip*):

(20)



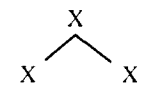
But there is no independent motivation for any of these solutions. They all weaken the independent phonetic definitions of prosodic domain types: (19) makes it difficult to give a consistent definition of utterance; (20) makes it difficult to give a consistent definition of intermediate phrase; and (18) posits an intermediate phonetic category for which there is no independent evidence – and a fortiori no phonetic definition. In all cases the justification for the intermediate layer of structure is entirely theory-internal – to avoid building a tree that violates the SLH.

8.2.4 *Compound prosodic domains*

If we wish to account for the evidence of fine differences of boundary strength, while at the same time keeping the number of prosodic domain types to a minimum of independently motivated categories, then we have no choice but to get rid of the SLH. The key to a more adequate theory of prosodic structure is to acknowledge the phonetic evidence for indeterminate depth of structure while at the same time preserving the intuition that prosodic structure is somehow flatter than the syntactic structure to which it corresponds. That is, I believe it is important to express the intuition that there is some essential difference between syntactic and prosodic structure; I take it that what we need to do is to weaken or relax the SLH, not (as I proposed in Ladd 1986) to abandon it altogether. The theoretical construct that makes this balancing act possible is what I have termed a Compound Prosodic Domain (CPD).

A CPD is a prosodic domain of a given type X whose immediate constituents *are themselves of type X*:

(21)



This is precisely the structure usually assumed for compounds in morphosyntax, and the analogy is intentional. A compound noun like *telephone call* is made up of two independently occurring nouns *telephone* and *call*, but at the same time

functions exactly like a single noun: the definition of noun is not compromised by the existence of such compounds. In the same way, [*A and B*] in (14) can be treated for certain purposes as a single *IP*, even though at the same time we recognise that it is composed of the two *IP*s *A* and *B*. As with morphosyntactic compounds, this analysis need not compromise the definitions of domain types: in particular, definitions in terms of boundaries can apply to simple and to compound domains in exactly the same way. That is, the right edges of both the simple *IP* [*B*] and the compound *IP* [*A and B*] are marked by a single boundary tone, as specified by Pierrehumbert and Beckman's definition of an *IP*. The fact that it is the very same boundary tone and the very same edge is of no consequence to the definitions. We may therefore diagram the prosodic structures corresponding to the two readings of (14) as in (16), that is, with a compound *IP* dominating either [*A and B*] or [*B and C*].

Strictly speaking, the notion of CPD involves recursive structure. If prosodic domains can be compound, then the depth of prosodic trees cannot be determinate, and this is clearly incompatible with the SLH. However, compounding is a limited kind of recursiveness. So long as the indeterminate depth of prosodic trees arises only through compounding and in no other way, then prosodic structure is still crucially different from syntactic structure, because the devices that create indeterminate depth in syntactic trees are richer and more powerful than compounding alone. There is no prosodic analogue to the sort of recursiveness seen in syntax whereby, for example, a sentence can contain noun phrases which in turn can contain sentences. Under the CPD proposal, there remains a clear sense in which phrases are 'higher' or more inclusive than feet, and feet are higher than syllables. *The categories of the prosodic hierarchy are still strictly ranked.* This means that, even if we allow CPDs, we are still able to express the 'flatness' of prosodic structure relative to syntax, and therefore do not lose the only important empirical advantage of the SLH.

On the other hand, if we allow CPDs, the empirical basis of the overall theory of prosodic phrasing is considerably strengthened. The notion of CPD allows us to accommodate decades of observations that have been forced into the straitjacket of 'big ProsPs versus little ProsPs', and to reconsider the early proposals by Liberman and Prince, Selkirk, and Nespor and Vogel on the ways in which prosodic domains can be related to each other. It almost certainly means that the number of prosodic domain types can be reduced rather than continually expanded: for example, Nespor and Vogel's 'clitic group' might be treated as a 'compound phonological word'. More relevantly for the topic of this book, we might also consider the possibility that the 'intermediate phrase' is nothing more than the smallest prosodic unit that can have a tune, and that the

distinction in type between the intermediate phrase and the intonational phrase can be eliminated.

In any case, such a reduction in the number of prosodic domain types would substantially strengthen the definitions of prosodic domains. It would allow us to identify any given boundary as being of one category or another on purely phonetic and phonological grounds, without as it were looking over our shoulder at the theoretical consequences for prosodic structure or for syntax–prosody mapping. For example, it would allow us to treat the boundary after *house* in (11) as being of the same type as the others (because it is marked by the same fall in pitch after the H* accent, and the same preboundary lengthening of the segments), but to recognise that it is also 'weaker' than the others. It could be weaker in two senses: first of all, it is abstractly weaker, by definition, in the sense that the two domains it separates (*that lay in the house* and *that Jack built*) together form a compound domain; but an obvious empirical hypothesis is that it is also weaker in some fairly literal phonetic sense for example, that it exhibits a shallower fall in pitch and/or less preboundary lengthening.

The task of developing and testing a full theory along the lines just sketched could easily occupy an entire monograph by itself, and I do not pursue the question further in detail. However, in the final section I consider some of the descriptive possibilities that arise from the view of prosodic structure sketched here.

8.3 Some descriptive consequences

In this section I briefly sketch three specific cases where recognising the centrality of prosodic structure in intonational phonology may help resolve some long-standing descriptive dilemmas. I presuppose some version of prosodic structure without the SLH, incorporating CPDs or something like them. I do not pretend that the brief sketches here provide complete answers to anything, but I offer them as pointers to potentially interesting and productive research questions.

8.3.1 Primary accent and focus

In section 7.1, I developed the idea that focus is signalled not by pitch accents in general, but by primary or nuclear accents, and suggested that primary accents in turn are merely indicative of the division of utterances into intermediate phrases, which is the true cross-linguistic basis of focus signalling. What I did not point out, though, is that there are plenty of cases where primary accent fails to signal focus. One such case involves the distinction between

'epithet' and 'literal' interpretations of words like *butcher* and *bastard*. In Ladd 1980a (p. 64ff.), I discussed a series of examples showing that one function of deaccenting is to favour the 'epithet' reading. While the distinction between the epithet and literal readings is not transparently a matter of focus, it seems clear that deaccenting and focus are part of the same cluster of phenomena, and I will therefore treat the literal/epithet distinction as equivalent to the focus/non-focus distinction for purposes of the point being made here. The basic phenomenon is seen in the following example:

(22) A: Everything OK after your operation?
B: Don't talk to me about it! I'd like to STRANGLE the butcher!

Deaccenting makes it possible to treat *butcher* in this exchange as an epithet referring to the doctor who performed the operation. Absence of deaccenting – that is, the sentence stress pattern that in most contexts would be interpreted as signalling neutral broad focus – turns B's contribution into an incoherent rant about a literal butcher, someone who sells meat:

(23) A: Everything OK after your operation?
B: Don't talk to me about it! I'd like to strangle the BUTCHER!

As I also noted in Ladd 1980a, the same distinction between epithet and literal meanings can be signalled by the *type* of accent, when the word in question is not sentence-final. With a secondary H* accent (what I originally called a 'B accent', following Bolinger), the noun can have the epithet reading:[3]

(24) A: Everything OK after your operation?
B: Don't talk to me about it!
H* H* H* LL%
The butcher charged me a thousand bucks!

With a primary accent (a Bolinger 'A accent'; in Pierrehumbert's terms an H* accent followed by an L phrase accent and an intermediate phrase boundary), *butcher* again refers to someone who sells meat:

(25) A: Everything OK after your operation?
B: Don't talk to me about it!
H* L H* H* LL%
The butcher charged me a thousand bucks!

3 I have taken this example directly from Ladd 1980a, complete with its seriously outdated assumptions about what counts as an outrageous medical expense.

So far, these observations are consistent with the idea that focus is signalled only by *primary* accent. In (24), we might say that B's entire sentence *The butcher charged me a thousand bucks* is a single intermediate phrase, and the accent on *butcher* is therefore secondary. In (25), on the other hand, *The butcher* is a separate intermediate phrase, so that *butcher* by definition bears a primary accent, which in turn means that it cannot be interpreted as deaccented.

This analysis breaks down, however, in a very emphatic rendition of the sentence. If we put primary accents on all four content words – that is, if we divide the sentence into four intermediate phrases – then the epithet interpretation of *butcher* is once again favoured:

(26) A: Everything OK after your operation?
B: Don't talk to me about it!
H* L H* L H* L H* LL%
The butcher charged me a thousand bucks!

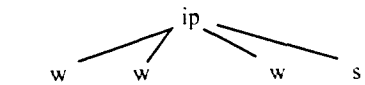
Here, *primary* accent is being used to signal the *epithet* reading. Intuitively, what seems to be happening is that in (25), *butcher* is separate enough from the prosody of the rest of the sentence that it counts as its own local maximum – this is essentially the same explanation as saying that it occurs as its own intermediate phrase – whereas in (24) and (26) the prominence of the accent on *butcher* is evaluated relative to the accents elsewhere in the sentence, and the local maximum for the sentence as a whole is on *bucks*.

This suggests that, for purposes of describing how focus is signalled, we must formulate our definition of 'primary' accent in a way that takes account of the discussion in the previous section about the indeterminate depth of prosodic structure. The intuitive explanation of the focus in the foregoing examples in terms of 'separateness' can be related to the notion of 'boundary strength': somehow the boundary between *butcher* and the rest of the sentence is 'stronger' in (25) than it is in (24) and (26). What all the epithet cases have in common is that *butcher* bears a degree of prominence that is in some sense *relatively less prominent* than some other accent in the utterance. That idea appears inconsistent with the observation that the boundary between *butcher* and the rest of the sentence is an intermediate phrase boundary in (25) and (26), but not in (24). However, if we consider the possibility of Compound Prosodic Domains, we may be able to salvage the intuitive explanation.

The traditional solution to this kind of problem would be to say that the phonetic cues to phrase boundaries are mysterious and elusive. For example, in order to salvage an account of focus based on primary accents, we might say that the valleys we identified as *ip* boundaries in (26) are not really *ip* boundaries,

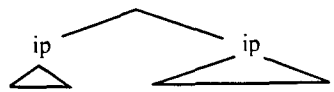
and therefore that the accents that fail to signal focus are not really primary. With the notion of CPD, we are not forced into the position of reasoning in this way. Instead, we simply treat each emphatic accent in (26) as the primary accent of a *local* intermediate phrase, but at the same time treat the entire sentence as a compound intermediate phrase in which *bucks* is the nucleus and *butcher* is prenuclear. That is, we would diagram examples (24)–(26) as follows:

(27) (corresponds to 24)



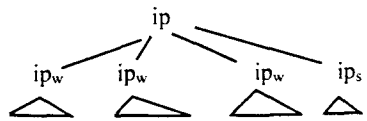
The butcher charged me a thousand bucks

(28) (corresponds to 25)



The butcher charged me a thousand bucks

(29) (corresponds to 26)



The butcher charged me a thousand bucks

This allows us both to maintain the definition of intermediate phrase and to express the differences of syntagmatic structure that seem to be involved in determining the interpretation of *butcher*. Obviously, if this proposal is to have real explanatory value, it will be necessary to constrain the range of circumstances in which we can posit compound intermediate phrases; ideally, we will be able to identify phonetic characteristics of compound domains as well as simple domains. The point here is to suggest a way forward, a basis for future research.

8.3.2 *Abstract tunes revisited: compound contours*

One of the most widely ignored problems in intonational phonology is the problem of tonal *dependencies* between prosodic phrases, such as the intonational dependency between main sentences and tags briefly illustrated in example

(9). Various writers have alluded to these problems (e.g. my own very brief discussion in Ladd 1980a: sect. 7.5); the most extensive treatment is Crystal's discussion of 'intonational subordination' (Crystal 1969: sect. 5.10, reprinted in Bolinger 1972a: 126–35), but there is no convincing account of how such dependencies work anywhere in the literature. As I argued in Ladd 1986, relaxing the Strict Layer Hypothesis makes it possible to consider a range of possible prosodic structures that may help to make sense of these dependencies.[4] Here I wish to discuss one case that seems to exemplify the relevance of CPDs to the understanding of tune–text association, namely the case of what I call *compound contours*.

It is not hard to demonstrate that 'compound contour' is a useful notion. Consider the contrast between the following two English contours, illustrated on a single short phrase:

(30) (a) L* H* LL%
Mary and Peter

(b) H* !H* LL%
Mary and Peter

The context might be a discussion of whom to invite as the fourth couple for a dinner party: (30a), with the 'surprise-redundancy contour' (Liberman 1975) suggests a sudden inspiration ('Why, I know who we can invite!'), while (30b), with downstepping of the nuclear accent on *Peter*, suggests rather the final decision of the person with the final say on the matter ('I say this is who we'll invite'). The same distinction applies to the following two utterances, each of which consists of two phrases:

(31) (a) L* H H* L L%
If Mary goes, Peter can go.

(b) H* H !H* L L%
If Mary goes, Peter can go.

Here the context might be a discussion of a children's outing of some sort. In (31a), the problem is that Peter is too young to go on his own; the speaker has had the sudden inspiration that Peter's big sister Mary could go, which in turn solves the problem ('Why, this makes it possible for Peter to go'). In (31b), the problem is rather that Peter thinks he is too old for whatever is being planned and does not want to go with his little sister Mary; the speaker – as

4 Some of the specific analyses in Ladd 1986 would not be permitted under the proposals outlined here, but the basic idea is the same.

in the previous example, someone with the final authority – is fed up with dealing with the siblings' disputes and makes a determination that Peter is to accompany Mary ('I say Peter should go'). If, at some level of description, we can treat the contours on (31a) and (31b) as being the same as those on (30a) and (30b), then this similarity of intonational nuance poses no theoretical problem. If, on the other hand, the phonology is prevented in principle from treating the contours in (31) as compound units that are identical to the simple units in (30), then the similarity of meaning is, in rigorous theoretical terms, accidental.

The significance of compound contours to these two problems of tune–text association is to show clearly that the abstract identity of the tune does not reside in the string of tones. If the surprise–redundancy tune can be manifested either as a sequence L* ... H*_L_L% (on a single phrase), or as a sequence L*_H ... H*_L_L% (on the two constituent phrases of a compound phrase), then the surface differences between the two sequences of tones must be seen as low-level differences conditioned by prosodic structure. That is, once we admit the notion of CPDs, and specifically compound tunes, we have further evidence that tonal specifications are abstract properties of *ips*, and that the surface string of accents and edge tones is a concrete and quite predictable consequence of a general process of tune–text association.

8.3.3 *'Metrical' pitch range effects*

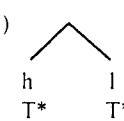
Finally, explicitly incorporating metrical structure into intonational phonology also seems likely to help us understand the phonology of pitch range. As we saw in section 5.2, much AM discussion of pitch range effects has revolved around the question of whether a given phenomenon is to be represented in the tonal string or treated as an 'extrinsic' scaling factor that modifies the phonetic realisation of the tone space. We specifically considered Sosa's critique of Pierrehumbert's reanalysis of the English 'hat pattern'. Recall that Beckman and Pierrehumbert (1986) replace Pierrehumbert's original description based on two different accent types (H* and H*+H) with an analysis expressed in terms of a single accent type affected by extrinsic factors ('high and compressed pitch range'). Sosa insists that we must recognise multiple accent types for this kind of distinction, on the grounds that 'the value of each tone in Hz is a function of its phonological characteristics ... in conjunction with the phonological characteristics of the preceding tone' and 'the tonal value of the [intervening] unaccented syllables is predictable by simple interpolation in the majority of cases' (1999: 107; my translation). Stated in this way, the issue seems like just another 'paralinguistic stalemate', of the sort discussed in section 1.4:

one analyst treats a given intonational phenomenon as linguistic and makes provision for it in the phonological analysis (in this case, the tonal string), while another analyst – or the same analyst a few years later – argues that the phenomenon is outside the system of linguistic contrasts and consequently should not be represented phonologically at all. The underlying question – paralinguistic versus linguistic, extrinsic versus intrinsic pitch range effects – appears in all cases to be the same.

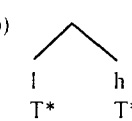
However, I believe that at least some of these stalemates can be resolved by recognising the relevance of prosodic structure to the control of pitch range. This can be seen if we consider once again the problem of downstep, which we have discussed extensively earlier in the book (sections 2.4, 3.1.3, 3.1.5). Recall that Pierrehumbert (1980) proposed that downstep is the result of local phonological rules operating on specific sequences of tones. Ladd (1983a) argued that this phonological analysis fails to express the fact that downstep in a language like English has a certain functional unity, regardless of the accent type to which it applies, and proposed a downstep 'feature' that can apply to any accent type and that has the effect of modifying the pitch range at specific points in the contour. In their rejoinder, Beckman and Pierrehumbert (1986) pointed out that such a downstep feature could apply not only to any accent type, but also to accents in any position, creating presumably meaningless tonal specifications in which the first accent in a series is downstepped (cf. chapter 3, footnote 5 on p. 99).

In order to deal with this problem, while at the same time preserving what I believe to be the correct assumption that downstep is a scaling factor orthogonal to accent type, I subsequently proposed (Ladd 1988, 1990b, and especially 1993b) that downstep involves a *syntagmatic relation* of pitch level between two accents or other prosodic constituents. This proposal, inspired by similar work on African tone languages by Clements (1983) and Huang (1980), is based directly on the notion of relative strength in metrical phonology. Just as two phonological constituents or domains can be related as either weak–strong or strong–weak (see section 2.2.2), so I proposed that at least certain kinds of prosodic constituents can also be related as either low–high or high–low. Thus:

(32) (a)



(b)



The relation in (32a) is that between an accent and a following downstepped accent; the relation in (32b) is that between an accent and a following non-downstepped one.[5]

The issue is ostensibly like the dispute between Sosa and Pierrehumbert about whether to treat downstep as part of the tonal string or as an extrinsic factor influencing the realisation of the tonal space. However, the 'metrical' interpretation of downstep gives us, in effect, the possibility of recognising not two but three kinds of factors involved in the scaling of tonal targets. These are:

- intrinsic – that is, different specifications relative to the tonal space (e.g. the difference between M tone and H tone)
- extrinsic – that is, overall modifications of the level and span of the tonal space (e.g. raising the voice for paralinguistic emphasis)

and, in addition

- metrical – that is, localisable linguistically meaningful modifications of the tonal space (e.g. downstep) controlled by relations in prosodic structure.

The third category is, as it were, somewhere 'in between' intrinsic and extrinsic factors – neither as categorically distinct as intrinsic linguistic factors that are phonologically encoded in the tonal string, nor as gradient and unformalisable as the paralinguistic effects that everyone agrees should be excluded from intonational phonology and treated as extrinsic. Such effects are not represented in the tonal string, but instead involve abstract relations between tones and between higher-level phonological constituents. Likely candidates for this type of description include not only downstep, but the *Anna/Manny* relation discovered by Liberman and Pierrehumbert 1984 (see section 2.3.3).

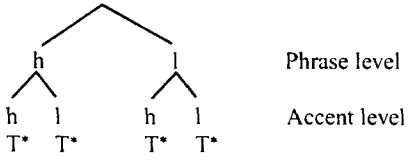
One obvious argument in favour of recognising an intermediate category of metrical pitch range effects is the fact that these effects can apparently be controlled independently of the clearly extrinsic global range modifications. That is, when speakers modify their overall range, the pitch range relations involved in downstep and in the *Anna/Manny* construction remain the same.

5 In order to clarify a point that has sometimes led to misunderstanding, I should emphasise that the relations h–l and l–h are not intended to supplant the relation w–s; that is, constituents may enter into relations of relative pitch level *in addition to* relations of relative prominence. The whole point of recognising downstep in a language like English (as we saw in section 2.4) is that a given accent may be 'nuclear' (and hence structurally more prominent than a preceding accent) even though it is 'downstepped' (and hence scaled at a lower pitch than the preceding accent). For more discussion of this point see Ladd 1989.

Indeed, if the metrical pitch range effects such as downstep and the *Anna/Manny* relations were not independent of overall paralinguistic modifications of pitch range, the device of getting speakers to raise their voice in order to study the metrical effects would not reveal anything systematic, and there would have been nothing to write about in sections 2.3.3 or 5.2.3! The very methodology presupposes that these effects can be controlled independently. Yet Beckman and Pierrehumbert (1992) suggest that the constant patterns in the experimental data discussed in those sections are merely the consequence of ill-defined 'speaker strategies' for making paralinguistic choices in the experimental situation: in this view it is only a remarkable coincidence that all these choices bear the same quantitative relation to one another whether the voice is raised or lowered.

The appropriateness of a metrical representation is further suggested by the fact that pitch-level relations can apply between phrases (and probably even larger constituents) as well as between individual accents. Several studies (e.g. Ladd 1988; van den Berg, Gussenhoven, and Rietveld 1992; Féry and Truckenbrodt 2005) have demonstrated that downstep can be 'nested'; that is, it is possible to have downstep within a series of short phrases, superimposed on an overall downward trend from phrase to phrase across the utterance as a whole. Truckenbrodt (2002, 2004) has presented a variety of evidence that not only downstep, but upstep and other register effects can be nested in the same way. Obviously, in a metrical analysis of pitch range effects, it is easy to express such nesting or embedding by extending the high–low relation shown in (32) to constituents at a higher level, as in the following:

(33)



That is, the metrical proposal makes it possible to treat nested register effects as a natural consequence of the phonological nature of downstep within phrases.

By allowing for pitch range relations to be nested like prominence relations, the analysis proposed here treats downstep as only one of a group of metrical or structural pitch range effects, in which pitch range shifts are used to signal syntactic and textual structure. For example, Ladd (1990a) presents preliminary evidence that the *Anna/Manny* relations can be nested in the same way as

downstep. The metrical analysis also makes sense of 'declination reset' – the upward modification of the pitch range at the beginning of a new stretch of declination – and 'final lowering' – the corresponding downward modification of the pitch range at the end. These are related to 'paragraph cues', whereby paragraphs, or other larger chunks of text, often begin quite high and end quite low (e.g. Lehiste 1975; Silverman 1987), and to discourse-level resettings of the sort noted by Brown, Currie, and Kenworthy (1980), whereby speakers tend to begin new topics high in their speaking range. In the view presented here, such resettings are simply the phonetic manifestation of phonological pitch range relations that can hold between prosodic constituents at different levels of structure.

An important aspect of my critique of Pierrehumbert's treatment of downstep is that it treats linguistically important pitch range effects in inconsistent ways. For Pierrehumbert, downstep within intermediate phrases is the phonetic effect of a phonological rule operating on certain specific sequences of accents. All other pitch range effects are explicitly classed as paralinguistic: for example, Beckman and Pierrehumbert (1986) claim that downward trends from phrase to phrase are paralinguistic, generated by a series of independent pitch range choices that merely 'mimic' downstep. If we recognise the existence of an intermediate metrical category of pitch range effects, we are not forced to adopt this view. Nor are we forced, with Sosa, to adopt a strictly initialising approach in which all scaling distinctions have to be represented in the tonal string. By proposing a category of pitch range effects that is intermediate between clearly intrinsic and clearly extrinsic, and by proposing a metrical representation of such pitch range relations, I am proposing that it is theoretically coherent to recognise the existence of factors that are *linguistic* but nevertheless *orthogonal to the tonal string*, including downstep, upstep at phrase boundaries, and the *Anna/Manny* relations.

Yet Beckman and Pierrehumbert obviously have a point. It is implausible to assume that the global paralinguistic effects on pitch range, and what I have called metrical pitch range effects, are completely unrelated. As Bolinger repeatedly pointed out (e.g. 1986: ch. 9), 'resetting' the pitch range at the beginning of a new topic is too obviously related to the paralinguistic function of raising the voice for the two to be completely separate systems. While I believe that the notion of metrical pitch range effects may go some way to clarifying the issues, the problem of the relationship between the linguistic and the paralinguistic still has not gone away. Perhaps a useful way of thinking about metrical effects on pitch range is that they represent a kind of phonologisation of the more basic paralinguistic functions.

The book ends, therefore, where it began: trying to untangle the linguistic and the paralinguistic. I believe that the three-way classification of pitch range effects proposed here provides the theoretical basis we need for designing empirical studies that will advance our understanding. I also believe, more generally, that the AM approach to pitch range as a problem of target scaling will make it possible to interpret the results of such studies in revealing ways. For now, we must stop short of drawing a clear boundary between language and paralanguage, and proposing the existence of an intermediate category is the best we can do. But the AM interpretation of the intermediate category as 'metrical' seems to hold out a great deal of promise.

References

Abbreviations

BLS	*Proceedings of the Berkeley Linguistics Society*
CLS	*Proceedings of the Chicago Linguistic Society*
ESCA	*European Speech Communication Association*
ICPhS	*Proceedings of the International Congress of Phonetic Sciences*
ICSLP	*Proceedings of the International Conference on Speech and Language Processing*
IEEE	Institute of Electrical and Electronics Engineers
IJAL	*International Journal of American Linguistics*
IPO	Institute for Perception Research
IULC	Indiana University Linguistics Club
JASA	*Journal of the Acoustical Society of America*
JIPA	*Journal of the International Phonetics Association*
JL	*Journal of Linguistics*
JML	*Journal of Memory and Language*
JPhon	*Journal of Phonetics*
Lg	*Language*
LgSp	*Language and Speech*
LI	*Linguistic Inquiry*
LVC	*Language Variation and Change*
MIT	Massachusetts Institute of Technology
NELS	*Proceedings of the North-East Linguistic Society*
NLLT	*Natural Language and Linguistic Theory*
SpComm	*Speech Communication*
ZAS	*Zentrum für Allgemeine Sprachwissenschaft*

Abe, Isamu. 1962. Call contours. *ICPhS* 7 (Helsinki): 519–23.

Abercrombie, David. 1991. *Fifty years in phonetics*. Edinburgh University Press.

Abramson, Arthur. 1962. The vowels and tones of Standard Thai: acoustical measurements and experiments. *IJAL* 28, no. 2, part 3.

Adriaens, L.M.H. 1991. *Ein Modell deutscher Intonation: eine experimentell-phonetische Untersuchung nach den perzeptiv relevanten Grundfrequenzänderungen in vorgelesenem Text*. PhD thesis, Eindhoven University of Technology.

Allerton, David and Alan Cruttenden. 1979. Three reasons for accenting a definite subject. *JL* 15: 49–53.

Altmann, Hans (ed.). 1988. *Intonationsforschungen*. Tübingen: Niemeyer.

Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson, and Regina Weinert. 1991. The HCRC map task corpus. *LgSp* 34: 351–66.

Anderson, Mark, Janet Pierrehumbert, and Mark Liberman. 1984. Synthesis by rule of English intonation patterns. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2.8.2–2.8.4.

Anderson, Stephen R. 1978. Tone features. In Fromkin 1978: pp. 133–75.

Anderson, Steven W. and William E. Cooper. 1986. Fundamental frequency patterns during spontaneous picture description. *JASA* 79: 1172–4.

Apple, W. and K. Hecht. 1982. Speaking emotionally: the relation between verbal and vocal communication of affect. *Journal of Personality and Social Psychology* 42: 864–975.

Árnason, Kristján. 2005. *Hljóð. Handbók um hljóðfræði og hljóðkerfisfræði*. Reykjavik: Almenna bókafélagið.

Arndt, W. 1960. 'Modal particles' in Russian and German. *Word* 16: 323–43.

Arvaniti, Amalia and Mary Baltazani. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In Jun 2005: pp. 84–117.

Arvaniti, Amalia and Gina Garding. 2007. Dialectal variation in the rising accents of American English. In J. Hualde and J. Cole (eds.), *Papers in laboratory phonology IX*. Berlin: Mouton de Gruyter, pp. 547–76.

Arvaniti, Amalia and Svetlana Godjevac. 2003. The origins and scope of final lowering in English and Greek. *ICPhS* 15 (Barcelona): 1077–80.

Arvaniti, Amalia and D. Robert Ladd. 1995. Tonal alignment and the representation of accentual targets. *ICPhS* 13 (Stockholm), vol. 4: 220–3.

Arvaniti, Amalia, D. Robert Ladd, and Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *JPhon* 26: 3–25.

2000. What is a starred tone? Evidence from Greek. In Broe and Pierrehumbert 2000: pp. 119–31.

2006. Phonetic effects of focus and 'tonal crowding' in intonation: evidence from Greek polar questions. *SpComm* 48: 667–96.

Ashby, Michael. 1978. A study of two English nuclear tones. *LgSp* 21: 326–36.

Atterer, Michaela and D. Robert Ladd. 2004. On the phonetics and phonology of 'segmental anchoring' of F_0: evidence from German. *JPhon* 32: 177–97.

Bard, Ellen Gurman, Anne H. Anderson, Catherine Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon, and Arlene Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *JML* 42: 1–22.

Bard, Ellen Gurman and Matthew Aylett. 1999. The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. *ICPhS* 14 (San Francisco): 1753–6.
Bard, Ellen Gurman, Daniel Robertson, and Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Lg* 72: 32–68.
Baumann, Stefan. 2006. *The intonation of givenness: evidence from German*. Tübingen: Niemeyer.
Beaugendre, Frédéric. 1994. *Une étude perceptive de l'intonation du français*. PhD thesis, Université de Paris XI.
Beaver, David, Brady Zack Clark, Edward Flemming, T. Florian Jaeger, and Maria Wolters. 2007. When semantics meets phonetics: acoustical studies of second-occurrence focus. *Lg* 83: 245–76.
Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
1995. Local shapes and global trends. *ICPhS* 13 (Stockholm), vol. 2: 100–7.
2007. ToBI home page. Online at www.ling.ohio-state.edu/~tobi (last accessed 22 March 2008). Linked to new English ToBI home page at www.tobihome.org (last accessed 22 March 2008).
Beckman, Mary E. and Gayle Ayers Elam. 1993 *Guidelines for ToBI labelling*. Linked to main ToBI website, online at www.tobihome.org (last accessed 29 October 2007).
Beckman, Mary E., Julia Hirschberg, and Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun 2005: pp. 9–54.
Beckman, Mary E. and John Kingston. 1990. Introduction. In Kingston and Beckman 1990: pp. 1–16.
Beckman, Mary E. and Janet B. Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3: 255–310.
1992. Comments on chapters 14 and 15. In Docherty and Ladd 1992: pp. 387–97.
Bing, Janet Mueller. 1979. *Aspects of English prosody*. PhD thesis, University of Massachusetts. Published 1985 by Garland Press, New York.
1980. The given/new distinction and the unmarked stress pattern. *NELS* XI (Ottawa): 13–21.
Birch, Stacy and Charles Clifton, Jr. 1995. Focus, accent, and argument structure: Effects on language comprehension. *LgSp* 38: 365–91.
Bishop, Judith and Janet Fletcher. 2005. Intonation in six dialects of Bininj Gun-Wok. In Jun 2005: pp. 331–61.
Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.
Bolinger, Dwight. 1951. Intonation: levels versus configurations. *Word* 7: 199–210. Reprinted in Bolinger 1965: pp. 3–16.
1954. English prosodic stress and Spanish sentence order. *Hispania* 37: pp. 152–6.
1958. A theory of pitch accent in English. *Word* 14: 109–49. Reprinted in Bolinger 1965: pp. 17–56.
1961a. *Generality, gradience, and the all-or-none*. The Hague: Mouton.

1961b. Contrastive accent and contrastive stress. *Lg* 37: 83–96. Reprinted in Bolinger 1965: pp. 101–17.
1964. Intonation: around the edge of language. *Harvard Educational Review* 34: 282–96. Reprinted (slightly abridged) in Bolinger 1972a: 19–29.
1965. *Forms of English: accent, morpheme, order*. Cambridge, MA: Harvard University Press.
(ed.). 1972a. *Intonation*. Harmondsworth: Penguin.
1972b. Accent is predictable (if you're a mind-reader). *Lg* 48: 633–44.
1978. Intonation across languages. In J. Greenberg (ed.), *Universals of human language*, vol. II: *Phonology*. Palo Alto, CA: Stanford University Press, pp. 471–524.
1982. Intonation and its parts. *Lg* 58: 505–33.
1986. *Intonation and its parts*. Palo Alto, CA: Stanford University Press.
1989. *Intonation and its uses*. Palo Alto, CA: Stanford University Press.
Bosch, Peter and Rob Van Der Sandt (eds.), 1999. *Focus: linguistic, cognitive, and computational perspectives*. Cambridge University Press.
Botinis, Antonis, Georgios Kouroupetroglou, and George Carayiannis (eds.), 1997. *Intonation: theory, models and applications* (Proceedings of an ESCA Workshop, Athens).
Braun, Bettina. 2006. Phonetics and phonology of thematic contrast in German. *LgSp* 49: 451–93.
Brazil, David, Malcolm Coulthard, and Catherine Johns. 1980. *Discourse intonation and language teaching*. London: Longman.
Bresnan, Joan. 1971. Sentence stress and syntactic transformations. *Lg* 47: 257–80.
Britain, David. 1992. Linguistic change in intonation: the use of high rising terminals in New Zealand English. *LVC* 4: 77–104.
Broe, Michael and Janet B. Pierrehumbert (eds.). 2000. *Papers in laboratory phonology V*. Cambridge University Press.
Brown, Gillian, Karen Currie, and Joanne Kenworthy. 1980. *Questions of intonation*. London: Croom Helm.
Browne, E.W. and James D. McCawley. 1965. Serbo-Croatian accent. In Erik C. Fudge (ed.), *Phonology*. Harmondsworth: Penguin, pp. 330–5.
Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.
1982. Developing the Swedish intonation model. *Working Papers 22* (Department of Linguistics and Phonetics, University of Lund) 51–116.
1987. How floating is focal accent? In K. Gregerson and H. Basbøll (eds.), *Nordic Prosody IV*. Odense University Press, pp. 41–9.
Bruce, Gösta and Eva Gårding. 1978. A prosodic typology for Swedish dialects. In E. Gårding, G. Bruce, and R. Bannert (eds.), *Nordic prosody*. Lund: Gleerup, pp. 219–28.
Brugos, Alejna, Stefanie Shattuck-Hufnagel, and Nanette Veilleux. 2006. *Transcribing Prosodic Structure of Spoken Utterances with ToBI* (MIT Open

Courseware). Linked to main ToBI website, online at www.tobihome.org (last accessed 22 March 2008).
Büring, Daniel. 1997. *The meaning of topic and focus: the 59th Street Bridge accent*. London: Routledge.
1999. Topic. In Bosch and van der Sandt 1999: pp. 142–65.
Calhoun, Sasha. 2006. *Information structure and the prosodic structure of English: a probabilistic relationship*. PhD thesis, University of Edinburgh.
Campbell, W.N. 1993. Automatic detection of prosodic boundaries in speech. *SpComm* 13: 343–54.
Campbell, W.N. and Mary E. Beckman. 1997. Stress, prominence and spectral tilt. In Botinis, Kouroupetroglou, and Carayiannis 1997: pp. 67–70.
Caspers, Johanneke and Vincent van Heuven. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–71.
Chafe, Wallace. 1974. Language and consciousness. *Lg* 50: 111–33.
1976. Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C. Li (ed.), *Subject and topic*. New York: Academic Press, pp. 25–56.
Chang, Nien-Chuang. 1958. Tones and intonation in the Chengtu dialect (Szechuan, China). *Phonetica* 2: 59–84. Reprinted in Bolinger 1972a: 391–413.
Chao, Y. R. 1932. A preliminary study of English intonation (with American variants) and its Chinese equivalents. T'sai Yuan Pei Anniversary Volume, supplementary vol. 1 of *Bulletin of the Institute of History and Philology of the Academica Sinica, Peiping*.
Chen, Aoju, Carlos Gussenhoven, and Toni Rietveld. 2004. Language-specificity in the perception of paralinguistic intonational meaning. *LgSp* 47: 311–49.
Chen, Matthew. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109–49.
Chen, Yiya and Yi Xu. 2006. Production of weak elements in speech – evidence from F_0 patterns of neutral tone in Standard Chinese. *Phonetica* 63: 47–75.
Ching, Marvin K.L. 1982. The question intonation in assertions. *American Speech* 57: 95–107.
Cho, Taehong and Peter Ladefoged. 1999. Variation and universals in VOT: evidence from 18 languages. *JPhon* 27: 207–29.
Chomsky, Noam. 1972. Deep structure, surface structure and semantic interpretation. In Noam Chomsky, *Studies on semantics in generative grammar*. The Hague: Mouton, pp. 62–119.
Chomsky, Noam and Morris Halle. 1968. *The sound pattern of English*. New York: Harper and Row.
Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *LI* 24: 239–97.
Clements, G.N. 1979. The description of terraced-level tone languages. *Lg* 55: 536–58.

1983. The hierarchical representation of tone features. In I. Dihoff (ed.), *Current approaches to African linguistics*, vol. I. Dordrecht: Foris.
1990. The status of register in intonation theory. In Kingston and Beckman 1990: pp. 58–71.
Clifton, Charles Jr and Lyn Frazier. 2004. Should given information come before new? Yes and no. *Memory and Cognition* 32: 886–95.
Cohen, Antonie, René Collier, and Johan 't Hart. 1982. Declination: construct or intrinsic feature of speech pitch? *Phonetica* 39: 254–73.
Cohen, Antonie and Johan 't Hart. 1967. On the anatomy of intonation. *Lingua* 19: 177–92.
Collier, René and Johan 't Hart. 1981. *Cursus Nederlandse Intonatie*. Louvain: Acco/De Horstink.
Connell, Bruce. 2000. The perception of lexical tone in Mambila. *LgSp* 43: 163–82.
Connell, Bruce and Amalia Arvaniti (eds.). 1995. *Papers in laboratory phonology IV*. Cambridge University Press.
Connell, Bruce and D. Robert Ladd. 1990. Aspects of pitch realisation in Yoruba. *Phonology* 7: 1–30.
Cooper, William, Stephen J. Eady, and Pamela R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *JASA* 77: 2142–55.
Cooper, William and Jeanne Paccia-Cooper. 1980. *Syntax and speech*. Cambridge, MA: Harvard University Press.
Cooper, William and John Sorensen. 1981. *Fundamental frequency in sentence production*. Heidelberg: Springer.
Couper-Kuhlen, Elizabeth and Margret Selting. 1996. *Prosody in conversation*. Cambridge University Press.
Croft, William. 1995. Intonation units and grammatical structure. *Linguistics* 33: 839–82.
Cruttenden, Alan. 1986. *Intonation*. Cambridge University Press.
1992a. Review of 't Hart, Collier, and Cohen 1990. *JL* 28: 522–6.
1992b. The origins of nucleus. *JIPA* 20: 1–9.
1993. The de-accenting and re-accenting of repeated lexical items. *Working Papers* 41 (Department of Linguistics and Phonetics, University of Lund): *Proceedings of the ESCA Workshop on Prosody, Lund, 27–29 September 1993*, 16–19.
1994a. 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, pp. 155–73.
1994b. *Gimson's Pronunciation of English*, fifth edition. London: Edward Arnold.
Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge University Press.
Culicover, Peter and Michael Rochemont. 1983. Stress and focus in English. *Lg* 59: 123–65.

Cutler, Anne, Delphine Dahan, and Wilma van Donselaar. 1997. Prosody in the comprehension of spoken language: a literature review. *LgSp* 40: 141–201.
Cutler, Anne and D. Robert Ladd (eds.). 1983. *Prosody: models and measurements*. Heidelberg: Springer.
Dainora, Audra. 2001. *An empirically based probabilistic model of intonation in American English*. PhD thesis, University of Chicago.
2006. Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. H. Whalen, and C. Best (eds.), *Laboratory phonology VIII*. Berlin: Mouton de Gruyter, pp. 107–32.
D'Alessandro, Christophe and Piet Mertens. 1995. Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language* 9: 257–88.
Dalton, Martha and Ailbhe Ní Chasaide. 2005. Tonal alignment in Irish dialects. *LgSp* 48: 441–64.
Daneš, Frantisek. 1967. Order of elements and sentence intonation. In *To honour Roman Jakobson: essays on the occasion of his seventieth birthday*. The Hague: Mouton, pp. 499–512. Reprinted in Bolinger 1972a: pp. 216–32.
Dawe, L.A., J.R. Platt, and R. J. Racine. 1993. Harmonic accents in inference of metrical structure and perception of rhythm patterns. *Perception & Psychophysics* 54: 794–807.
Delattre, Pierre. 1963. Comparing the phonetic features of English, German, Spanish and French. *International Review of Applied Linguistics* 1: 193–210. Reprinted (slightly revised) 1965, as chapter 2 of *Comparing the phonetic features of English, German, Spanish and French*. Heidelberg: Julius Groos, pp. 23–43.
1966. Les dix intonations de base du français. *French Review* 40: 1–14.
Dell, François. 1984. L'accentuation dans les phrases en français. In F. Dell, J.-R. Vergnaud, and D. Hirst (eds.), *Les représentations en phonologie*. Paris: Hermann, pp. 65–112.
de Pijper, Jan Roelof. 1983. *Modelling British English intonation*. Dordrecht: Foris.
de Pijper, Jan Roelof and Angelien Sanderman. 1995. On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *JASA* 96: 2037–47.
de Sivers, Fanny. 1965. L'unité intonationelle d'interrogation en hongrois. *La Linguistique* 1: 75–112.
Dilley, Laura. 2005. *The phonetics and phonology of tonal systems*. PhD thesis, MIT.
Dilley, Laura and Meredith Brown. 2007. Effects of pitch range variation on F_0 extrema in an imitation task. *JPhon* 35: 523–51.
Dilley, Laura, D. Robert Ladd, and Astrid Schepman. 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *JPhon* 33: 115–19.
D'Imperio, Mariapaola. 2002. Italian intonation: An overview and some questions. *Probus* 14: 37–69.

D'Imperio, Mariapaola and David House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of Eurospeech* (Rhodes), pp. 251–4.
Docherty, Gerard and D. Robert Ladd (eds.). 1992. *Papers in laboratory phonology II*. Cambridge University Press.
Donohue, Mark. 1997. Tone systems in New Guinea. *Linguistic Typology* 1: 347–86.
Downing, Laura J. 2008. Focus and prominence in Chichewa, Chitumbuka, and Durban Zulu. *ZAS Papers in Linguistics (ZASPiL)* 49: 47–65.
Dresher, B. Elan. 1994. The prosodic basis of the Tiberian Hebrew system of accents. *Lg* 70: 1–52.
Duanmu, San. 1994. Against contour tone units. *LI* 25: 555–608.
Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastián, and Jacques Mehler. 1997. A destressing 'deafness' in French? *JML* 36: 406–21.
Dupoux, Emmanuel, Sharon Peperkamp, and Nuria Sebastián-Gallés. 2001. A robust method to study stress 'deafness'. *JASA* 110: 1606–18.
Eady, S.J. and W.E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *JASA* 80: 402–15.
Eady, S.J., W.E. Cooper, G.V. Klouda, P.R. Mueller, and D.W. Lotts. 1986. Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments. *LgSp* 29: 233–51.
Earle, M.A. 1975. *An acoustic phonetic study of Northern Vietnamese tones*. Santa Barbara: Speech Communications Research Laboratory.
Ebing, E.F. 1994. Towards an inventory of perceptually relevant pitch movements for Indonesian. In Odé and van Heuven 1994: pp. 181–210.
Elordieta, Gorka and Nagore Calleja. 2005. Microvariation in accentual alignment in Basque Spanish. *LgSp* 48: 397–439.
Elordieta, Gorka, Sónia Frota, Pilar Prieto, and Marina Vigário. 2003. Effects of constituent length and syntactic branching on intonational phrasing in Ibero-Romance. *ICPhS 15* (Barcelona): 487–90.
Erteschik-Shir, Nomi. 2007. *Information structure: the syntax-discourse interface*. Oxford University Press.
Faber, David. 1987. The accentuation of intransitive sentences in English. *JL* 23: 341–58.
Face, Timothy L. 2002. *Intonational marking of contrastive focus in Castilian Spanish*. Munich: Lincom Europa.
Fant, Gunnar and Anita Kruckenberg. 1994. Notes on stress and word accent in Swedish. *Speech Transmission Laboratory Quarterly Status Report*, 15 October 1994. Department of Speech Communication and Music Acoustics, Royal Institute of Technology, Stockholm.
Féry, Caroline. 1993. *German intonational patterns*. Tübingen: Niemeyer.
Féry, Caroline and Vieri Samek-Lodovici. 2006. Focus projection and prosodic prominence in nested foci. *Lg* 82: 131–50.
Féry, Caroline and Hubert Truckenbrodt. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59: 223–43.

Fletcher, Janet and Jonathan Harrington. 2001. High-Rising Terminals and Fall-Rise Tunes in Australian English. *Phonetica* 58: 215–29.
Fónagy, I., J. Fónagy, and J. Sap. 1979. A la recherche de traits pertinents prosodiques du français parisien. *Phonetica* 36: 1–20.
Fowler, Carol A. and J. Housum. 1987. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *JML* 26: 489–504.
Fromkin, Victoria (ed.). 1978. *Tone*. New York: Academic Press.
Frota, Sónia. 1998. *Prosody and focus in European Portuguese*. PhD thesis, University of Lisbon.
2002. Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation. *Probus* 14: 113–46.
Fry, D.B. 1955. Duration and intensity as physical correlates of linguistic stress. *JASA* 27: 765–8.
1958. Experiments in the perception of stress. *LgSp* 1: 126–52.
Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.), *The production of speech*. Heidelberg: Springer-Verlag, pp. 39–55.
Fujisaki, Hiroya and Keikichi Hirose. 1982. Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In *Preprints of Papers. Working Group on Intonation. Thirteenth International Congress of Linguists, Tokyo*, pp. 57–70.
Gandour, Jackson T. 1978. The perception of tone. In Fromkin 1978: pp. 41–76.
Gårding, Eva. 1983. A generative model of intonation. In Cutler and Ladd 1983: pp. 11–25.
1987. Speech act and tonal pattern in standard Chinese – constancy and variation. *Phonetica* 44: 13–29.
Gay, Thomas. 1978. Physiological and acoustic correlates of perceived stress. *LgSp* 21: 347–53.
Gee, J.P. and François Grosjean. 1983. Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15: 411–58.
German, James, Janet Pierrehumbert, and Stefan Kaufmann. 2006. Evidence for phonological constraints on nuclear accent placement. *Lg* 82: 151–68.
Gibbon, Dafydd. 1976. *Perspectives of intonation analysis*. Bern: Lang.
1998. Intonation in German. In Hirst and di Cristo 1998: pp. 78–95.
Giegerich, Heinz. 1985. *Metrical phonology and phonological structure: German and English*. Cambridge University Press.
1992. *English phonology: an introduction*. Cambridge University Press.
Glasberg, B.R. and Brian Moore. 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research* 47: 103–38.
Goldsmith, John. 1976. *Autosegmental phonology*. PhD thesis, MIT. Distributed by IULC and published 1979 by Garland Press, New York.
Gordon, Matthew K. 2005. Intonational phonology of Chickasaw. In Jun 2005: pp. 301–30.
Gorman, James. 1993. On language. *New York Times Magazine*, 20 August.

Gósy, Mária and Jacques Terken. 1994. Question marking in Hungarian: timing and height of pitch peaks. *JPhon* 22: 269–81.
Grabe, Esther. 1998. *Comparative intonational phonology: English and German*. PhD thesis, Max Planck Institute for Psycholinguistics, Nijmegen.
2003. *English Intonation in the British Isles: The IViE corpus*. Online at www.phon.ox.ac.uk/old_IViE (last accessed 22 March 2008).
Grabe, Esther, Brechtje Post, Francis Nolan, and Kimberley Farrar. 2000. Pitch accent realization in four varieties of British English. *JPhon* 28: 161–85.
Grice, Martine. 1995a. *The intonation of interrogation in Palermo Italian: implications for intonation theory*. Tübingen: Niemeyer.
1995b. Leading tones and downstep in English. *Phonology* 12: 183–233.
Grice, Martine, Stefan Baumann, and Ralf Benzmüller. 2005. German intonation in autosegmental-metrical phonology. In Jun 2005: pp. 55–83.
2007. *GToBI*. Online at www.uni-koeln.de/phil-fak/phonetik/gtobi/index.html (last accessed 22 March 2008).
Grice, Martine, Mariapaola D'Imperio, Michelina Savino, and Cinzia Avesani. 2005. Strategies for intonation labelling across varieties of Italian. In Jun 2005: pp. 362–89.
Grice, Martine, D. Robert Ladd, and Amalia Arvaniti (eds.). 2000. On the place of phrase accents in intonational phonology. *Phonology* 17: 143–85.
Grice, Martine, Matthias Reyelt, Ralf Benzmüller, J. Mayer, and Anton Batliner. 1996. Consistency in transcription and labelling of German intonation with GToBI. *ICSLP 4* (Philadelphia): 1716–19.
Grønnum, Nina. *See also* Thorsen.
1991. Prosodic parameters in a variety of regional Danish standard languages, with a view towards Swedish and German. *Phonetica* 47: 188–214.
Grosz, Barbara and Candy Sidner. 1986. Attention, intention, and the structure of discourse. *Journal of Computational Linguistics* 12: 175–204.
Grundstrom, Allan W. 1973. L'intonation des questions en français standard. In A. Grundstrom and P. Léon (eds.), *Interrogation et intonation en français standard et français canadien*. Montreal: Didier, pp. 19–51.
Gumperz, John. 1982. *Discourse strategies*. Cambridge University Press.
Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Lg* 69: 274–307.
Gussenhoven, Carlos. 1983a. Focus, mode, and the nucleus. *JL* 19: 377–417.
1983b. Testing the reality of focus domains. *LgSp* 26: 61–80.
1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
1993. The Dutch foot and the chanted call. *JL* 29: 37–63.
2000. The boundary tones are coming: On the nonperipheral realisation of boundary tones. In Broe and Pierrehumbert 2000: pp. 132–51.
2004. *The phonology of tone and intonation*. Cambridge University Press.
2005. Transcription of Dutch intonation. In Jun 2005: pp. 118–45.

Gussenhoven, Carlos, Bruno H. Repp, A.C.M. Rietveld, H.H. Rump, and Jacques Terken. 1997. The perceptual prominence of fundamental frequency peaks. *JASA* 102: 3009–22.

Gussenhoven, Carlos and Toni Rietveld. 1988. Fundamental frequency declination in Dutch: testing three hypotheses. *JPhon* 16: 355–69.

1991. An experimental evaluation of two nuclear-tone taxonomies. *Linguistics* 29: 423–49.

Gussenhoven, Carlos, Toni Rietveld, Joop Kerkhoff, and Jacques Terken. 2003. *Transcription of Dutch Intonation (ToDI)*, second edition. Online at http://todi.let.kun.nl/ToDI/home.htm (last accessed 22 March 2008).

Gussenhoven, Carlos and Peter van der Vliet. 1999. The phonology of tone and intonation in the Dutch dialect of Venlo. *JL* 35: 99–135.

Haan, Judith. 2001. *Speaking of questions: an exploration of Dutch question intonation*. Utrecht: LOT.

Hadding-Koch, Kerstin. 1961. *Acoustico-phonetic studies in the intonation of southern Swedish*. Lund: Gleerup.

Hadding-Koch, Kerstin and Michael Studdert-Kennedy. 1964. An experimental study of some intonation contours. *Phonetica* 11: 175–85. Reprinted in Bolinger 1972a: pp. 348–58.

Haggo, Douglas C. 1987. *The structure of English tonal morphemes*. PhD thesis, University of Canterbury (NZ).

Halle, Morris and Jean-Roger Vergnaud. 1987. *An essay on stress*. Cambridge, MA: MIT Press.

Halliday, M.A.K. 1960. Categories of the theory of grammar. *Word* 17: 241–92.

1966. The concept of rank: a reply. *JL* 2: 110–18.

1967a. *Intonation and grammar in British English*. The Hague: Mouton.

1967b. Notes on transitivity and theme in English (part II). *JL* 3: 199–244.

1970. *A course in spoken English: intonation*. Oxford University Press.

Halliday, M.A.K. and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Hannon, E.E., J.S. Snyder, T. Eerola, and C.L. Krumhansl. 2004. The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance* 30: 956–74.

Haspelmath, Martin. 1997. *Indefinite pronouns*. Oxford University Press.

2007. Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology* 11: 119–32.

Hatcher, Anna G. 1956. Theme and underlying question: two studies in Spanish word order. *Word* 12 (Monograph no. 3).

Haugen, Einar and Martin Joos. 1952. Tone and intonation in East Norwegian. *Acta Philologica Scandinavica* 22: 41–64. Reprinted in Bolinger 1972a: pp. 414–36.

Hawkins, Sarah and Paul Warren. 1991. Factors affecting the given-new distinction in speech. *ICPhS* 12 (Aix-en-Provence), vol. 4: 66–9.

Hayes, Bruce. 1981. *A metrical theory of stress rules*. PhD thesis, MIT. Distributed by IULC.

1995. *Metrical stress theory: principles and case studies*. University of Chicago Press.

Hayes, Bruce and Aditi Lahiri. 1991. Bengali intonational phonology. *NLLT* 9: 47–96.

Heine, Bernd and Derek Nurse (eds.). 2000. *African languages: an introduction*. Cambridge University Press.

Hellmuth, Samantha. 2007. The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review* 24: 291–316.

Henderson, Eugénie. 1952. The main features of Cambodian pronunciation. *Bulletin of the School of Oriental and African Studies* 14: 149–74.

Herman, Rebecca. 1996. Final lowering in Kipare. *Phonology* 13: 171–96.

Hermes, Dik. 2006. Stylization of pitch contours. In S. Sudhoff *et al.* (eds.), *Methods in empirical prosody research*. Berlin: Walter de Gruyter, pp. 29–61.

Hermes, Dik and H.H. Rump. 1994. Perception of prominence in speech intonation induced by rising and falling pitch movements. *JASA* 96: 83–92.

Hermes, Dik and Joost van Gestel. 1991. The frequency scale of speech intonation. *JASA* 90: 97–102.

Himmelmann, Nikolaus. Unpublished ms. Notes on Waima'a intonation.

Hirschberg, Julia and Janet Pierrehumbert. 1986. Intonational structuring of discourse. *Proceedings of the 24th Association for Computational Linguistics, New York*, pp. 136–44.

Hirschberg, Julia and Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. *JPhon* 20: 241–51.

1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24: 407–12.

Hirst, Daniel. 1983. Structures and categories in prosodic representations. In Cutler and Ladd 1983: pp. 93–109.

Hirst, Daniel and Albert Di Cristo (eds.). 1998. *Intonation systems: A survey of twenty languages*. Cambridge University Press.

Hirst, Daniel, Albert Di Cristo, and Robert Espesser. 2000. Levels of representation and levels of analysis for the description of intonation systems. In Horne 2000: pp. 51–88.

Hobbs, Jerry. 1990. The Pierrehumbert–Hirschberg theory of intonational meaning made simple: comments on Pierrehumbert and Hirschberg. In P.R. Cohen, J. Morgan, and M.E. Pollack (eds.), *Intentions in communication*. Cambridge, MA: MIT Press, pp. 313–23.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore: Waverley Press.

1958. *A course in modern linguistics*. New York: Macmillan.

Horne, Merle (ed.). 2000. *Prosody: theory and experiment (Studies presented to Gösta Bruce)*. Dordrecht: Kluwer.

Horvath, Julia. 1986. *Focus in the theory of grammar and the syntax of Hungarian*. Dordrecht: Foris.
Hualde, José Ignacio. 2002. Intonation in Spanish and other Ibero-Romance languages: overview and status quaestionis. In C. Wiltshire and J. Camps (eds.), *Romance phonology and variation. Selected papers from the 30th Linguistic Symposium on Romance Languages, Gainesville, Florida, February 2000*. Amsterdam: Benjamins, pp. 101–15.
Huang, C.-T. James. 1980. The metrical structure of terraced-level tones. *NELS 11* (Ottawa): 257–70.
Huddleston, Rodney. 1965. Rank and depth. *Lg* 41: 574–86. Reprinted in M.A.K. Halliday and J.R. Martin (eds.), *Readings in systemic linguistics*. London: Batsford.
Huss, Volker. 1978. English word stress in the postnuclear position. *Phonetica* 35: 86–105.
Hyman, Larry. 2006. Word-prosodic typology. *Phonology* 23: 225–57.
Hyman, Larry, Francis Katamba, and Livingstone Walusimbi. 1987. Luganda and the strict layer hypothesis. *Phonology Yearbook* 4: 87–108.
Inkelas, Sharon and William Leben. 1990. Where phonology and phonetics intersect: the case of Hausa intonation. In Kingston and Beckman 1990: pp. 17–34.
Inkelas, Sharon and Draga Zec. 1988. Serbo-Croatian pitch accent. *Lg* 64: 227–48.
Isačenko, A. and H.-J. Schädlich. 1970. *A model of Standard German intonation*. The Hague: Mouton.
Ishihara, Takeshi. 2003. A phonological effect on tonal alignment in Tokyo Japanese. *ICPhS* 15 (Barcelona): 615–19.
Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
Jacobs, Joachim. 1982. Neutraler und nicht-neutraler Satzakzent im Deutschen. In Th. Vennemann (ed.), *Silben, Segmente, Akzente*. Tübingen: Niemeyer, pp. 141–69.
1997. I-Topikalisierung. *Linguistische Berichte* 168: 91–133.
Jakobson, Roman, Gunnar Fant, and Morris Halle. 1952. *Preliminaries to speech analysis*. Cambridge, MA: MIT Press.
Jakobson, Roman and Morris Halle. 1971. *Fundamentals of language*, second edition. The Hague: Mouton.
Joos, Martin (ed.). 1957. *Readings in linguistics*, vol. I. University of Chicago Press.
Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody*. PhD thesis, Ohio State University. Published 1996 by Garland Press, New York.
(ed.). 2005. *Prosodic typology: the phonology of intonation and phrasing*. Oxford University Press.
Kager, René. 1999. *Optimality theory*. Cambridge University Press.

Kanerva, Jonni. 1990. Focusing on phonological phrases in Chichewa. In S. Inkelas and D. Zec (eds.), *The phonology-syntax connection*. University of Chicago Press, pp. 145–61.
Keating, Patricia (ed.). 1993. *Papers in laboratory phonology* III. Cambridge University Press.
Kefer, Michel. 1986. Some sentence accent universals referring to noun phrases. Unpublished paper, University of Liège.
Kendon, Adam. 2004. *Gesture: visible action as utterance*. Cambridge University Press.
Kiefer, Ferenc. 1967. *On emphasis and word order in Hungarian*. The Hague: Mouton.
Kim, A.H. 1988. Preverbal focusing and type XXIII languages. In M. Hammond, E. Moravcsik and J.R. Wirth (eds.), *Studies in syntactic typology*. Amsterdam: Benjamins, pp. 147–69.
King, Heather B. 1998. *The declarative intonation of Dyirbal: an acoustic analysis*. Munich: Lincom Europa.
Kingdon, Roger. 1958. *The groundwork of English intonation*. London: Longman.
Kingston, John and Mary E. Beckman (eds.). 1990. *Papers in laboratory phonology* I. Cambridge University Press.
Kiparsky, Paul. 1966. Über den deutschen Akzent. *Studia Grammatica* 7: 69–98.
Kiss, Katalin É. 1998. Identificational focus versus information focus. *Lg* 74: 245–73.
Knight, Rachael-Anne and Francis Nolan. 2006. The effect of pitch span on intonational plateaux. *JIPA* 36: 21–38.
Knowles, G.O. 1973. *Scouse: the urban dialect of Liverpool*. PhD thesis, University of Leeds.
Kochanski, Greg and Chilin Shih. 2003. Prosody modeling with soft templates. *SpComm* 39: 311–52.
Kochanski, Greg, Chilin Shih and Hongyan Jing. 2003. Hierarchical structure and word strength prediction of Mandarin prosody. *International Journal of Speech Technology* 6: 33–43.
Kohler, Klaus J. 1987. Categorical pitch perception. *ICPhS* 11 (Tallinn), vol. 5: 331–3.
1990. Macro and micro F_0 in the synthesis of intonation. In Kingston and Beckman 1990: pp. 115–38.
2005. Timing and communicative functions of pitch contours. *Phonetica* 62: 88–105.
König, Ekkehard. 1991. *The meaning of focus particles: a comparative perspective*. London: Routledge.
Koopmans-van Beinum, F.J. and D.R. van Bergem, 1989. The role of 'given' and 'new' in the production and perception of vowel contrasts in read text and in spontaneous speech. *Proceedings of Eurospeech* (Paris), pp. 113–16.

Kori, S. 1987. The tonal behavior of Osaka Japanese: an interim report. *Working Papers in Linguistics* 36 (Ohio State University): 31–61.
Kornai, András and Geoffrey K. Pullum. 1990. The X-bar theory of phrase structure. *Lg* 66: 24–50.
Kramer, E. 1964. Elimination of verbal cues in judgments of emotion from voice. *Journal of Abnormal and Social Psychology* 68: 390–6.
Krifka, Manfred. 1991. A compositional semantics for multiple focus constructions. In S. Moore and A. Wyner (eds.), *Proceedings of SALT I, Cornell*, pp. 127–58.
2006. Association with focus phrases. In V. Molnár and S. Winkler (eds.), *The architecture of focus*. Berlin: Mouton de Gruyter, pp. 105–36.
Kruijff-Korbayová, Ivana and Mark Steedman. 2003. Discourse and information structure. *Journal of Logic, Language and Information* 12: 249–59.
Kruyt, J.G. 1985. *Accents from speakers to listeners. An experimental study of the production and perception of accent patterns in Dutch*. PhD thesis, University of Leiden.
Kubozono, Haruo. 1987. *The organization of Japanese prosody*. PhD thesis, University of Edinburgh. Published 1993 by Kurishio Publishers, Tokyo.
1989. Syntactic and rhythmic effects on downstep in Japanese. *Phonology* 6: 39–67.
Ladd, D. Robert. 1978. Stylized intonation. *Lg* 54: 517–39.
1980a. *The structure of intonational meaning: evidence from English*. Bloomington: Indiana University Press.
1980b. English compound stress. *NELS* 10. Reprinted 1984 in D. Gibbon and H. Richter (eds.), *Intonation, accent and rhythm*. Berlin: Walter de Gruyter, pp. 253–66.
1983a. Phonological features of intonational peaks. *Lg* 59: 721–59.
1983b. Levels versus configurations, revisited. In F.B. Agard, G.B. Kelley, A. Makkai, and V.B. Makkai (eds.), *Essays in honor of Charles F. Hockett*. Leiden: E. J. Brill, pp. 93–131.
1984. Declination: a review and some hypotheses. *Phonology Yearbook* 1: 53–74.
1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3: 311–40.
1987a. Review of Bolinger 1986. *Lg* 63: 637–43.
1987b. A phonological model of intonation for use in speech synthesis by rule. In *Proceedings of the European Conference on Speech Technology, Edinburgh*, pp. 21–4.
1988. Declination 'reset' and the hierarchical organization of utterances. *JASA* 84: 530–44.
1989. Review of Pierrehumbert and Beckman 1988. *JL* 25: 519–26.
1990a. Intonation: emotion vs. grammar. (Review of Bolinger 1989.) *Lg* 66: 806–16.
1990b. Metrical representation of pitch register. In Kingston and Beckman 1990: pp. 35–57.

1992. An introduction to intonational phonology. In Docherty and Ladd 1992: pp. 321–34.
1993a. In defense of a metrical theory of intonational downstep. In H. v. d. Hulst and K. Snider (eds.), *The representation of tonal register*. Dordrecht: Foris, pp. 109–32.
1993b. Constraints on the gradient variability of pitch range (or) Pitch Level 4 lives! In Keating 1993: pp. 43–63.
1993c. On the theoretical status of 'the baseline' in modelling intonation. *LgSp* 36: 435–51.
1996. *Intonational phonology*, first edition. Cambridge University Press.
2001. Intonation. In M. Haspelmath, E. König, W. Oesterreicher, W. Raible (eds.), *Language typology and language universals: an international handbook*, vol. 2. Berlin: Walter de Gruyter, pp. 1380–90.
2006. Segmental anchoring of pitch movements: autosegmental association or gestural coordination? *Rivista di linguistica [Italian Journal of Linguistics]* 18: 19–38.
Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner, and Astrid Schepman. 1999. Constant 'segmental anchoring' of F_0 movements under changes in speech rate. *JASA* 106: 1543–54.
Ladd, D. Robert, Ineke Mennen, and Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *JASA* 107: 2685–96.
Ladd, D. Robert and Rachel Morton. 1997. The perception of intonational emphasis: continuous or categorical? *JPhon* 25: 313–42.
Ladd, D. Robert and Astrid Schepman. 2003. 'Sagging transitions' between high pitch accents in English: experimental evidence. *JPhon* 31: 81–112.
Ladd, D. Robert, Astrid Schepman, Laurence White, Louise May Quarmby, and Rebekah Stackhouse. Forthcoming. Phonological factors in F_0 alignment: evidence from Scottish Standard English. Submitted to *JPhon*.
Ladd, D. Robert, Klaus R. Scherer, and Kim Silverman. 1986. An integrated approach to studying intonation and attitude. In C. Johns-Lewis (ed.), *Intonation in discourse*. London: Croom Helm, pp. 125–38.
Ladd, D. Robert, Kim Silverman, Frank Tolkmitt, Günther Bergmann, and Klaus R. Scherer. 1985. Evidence for the independent function of intonation contour type, voice quality and F_0 range in signalling speaker affect. *JASA* 78: 435–44.
Ladd, D. Robert and Jacques Terken. 1995. Modelling intra- and inter-speaker pitch range variation. *ICPhS* 13 (Stockholm), vol. 2: 386–9.
Ladefoged, Peter. 1967. Linguistic phonetics. *UCLA Working Papers in Phonetics* 6.
1982. *A course in phonetics*, second edition. New York: Harcourt Brace Jovanovich.

1983. The linguistic use of different phonation types. In D. Bless and J. Abbs (eds.), *Vocal fold physiology: contemporary research and clinical issues*. San Diego: College Hill Press, pp. 351–60.
Ladefoged, Peter and D.E. Broadbent. 1957. Information conveyed by vowels. *JASA* 29: 98–104.
Laksman, Myrna. 1994. Location of stress in Indonesian words and sentences. In Odé and van Heuven 1994: pp. 108–39.
Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge University Press.
Langendoen, D. Terence. 1975. Finite-state parsing of phrase structure languages and the status of readjustment rules in the grammar. *LI* 6: 533–54.
Laniran, Yetunde O. 1992. *Intonation in tone languages: the phonetic implementation of tones in Yoruba*. PhD thesis, Cornell University.
Laniran, Yetunde O. and G.N. Clements. 2003. Downstep and high raising: interacting factors in Yoruba tone production. *JPhon* 31: 203–50.
Lass, Roger. 1984. Vowel system universals and typology: prologue to theory. *Phonology Yearbook* 1: 75–111.
LaVelle, C.R. 1974. An experimental study of Yoruba tone. *Studies in African Linguistics*, supplement 5: 185–94.
Laver, John. 1980. *The phonetic description of voice quality*. Cambridge University Press.
1994. *Principles of phonetics*. Cambridge University Press.
Leben, William. 1973. *Suprasegmental phonology*. PhD thesis, MIT. Published 1980 by Garland Press, New York.
1976. The tones in English intonation. *Linguistic Analysis* 2: 69–107.
Lee, W.R. 1956. Fall–rise intonations in English. *English Studies* 37: 62–72.
Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: MIT Press.
1975. The phonetic structure of paragraphs. In Antonie Cohen and S.G. Nooteboom (eds.), *Structure and process in speech perception*. Cambridge, MA: Springer, pp. 195–203.
Lehiste, Ilse and Pavle Ivić. 1980. The intonation of yes-and-no questions: a new Balkanism? *Balkanistica* 6: 45–53.
1986. *Word and sentence prosody in Serbo-Croatian*. Cambridge, MA: MIT Press.
Levi, Susannah. 2005. Acoustic correlates of lexical accent in Turkish. *JIPA* 35: 73–97.
Levin, Beth and M. Rappaport-Hovav. 1995. *Unaccusativity: at the syntax–lexical semantics interface*. Cambridge, MA: MIT Press.
Levy, Elena T. and David McNeill. 1992. Speech, gesture, and discourse. *Discourse Processes* 15: 277–301.
Liberman, Mark. 1975. *The intonational system of English*. PhD thesis, MIT. Distributed 1978 by IULC.

Liberman, Mark. 2007. How about the Germans? Posting to *Language Log*, 14 November 2007. Online at http://itre.cis.upenn.edu/~myl/languagelog/archives/005113.html (last accessed 22 March 2008).
Liberman, Mark and Janet Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oerhle (eds.), *Language sound structure*. Cambridge, MA: MIT Press, pp. 157–233.
Liberman, Mark and Alan Prince. 1977. On stress and linguistic rhythm. *LI* 8: 249–336.
Liberman, Mark and Ivan Sag. 1974. Prosodic form and discourse function. *CLS* 10: 416–27.
Liberman, Mark, J. Michael Schultz, Soonhyun Hong, and Vincent Okeke. 1993. The phonetic interpretation of tone in Igbo. *Phonetica* 50: 147–60.
Lickley, Robin, Astrid Schepman, and D. Robert Ladd. 2005. Alignment of 'phrase accent' low in Dutch falling rising questions: theoretical and methodological implications. *LgSp* 48: 157–83.
Liddell, Scott. 1977. *An investigation into the syntactic structure of American Sign Language*. PhD thesis, University of California, San Diego.
Lieberman, Philip. 1960. Some acoustic correlates of word stress in American English. *JASA* 32: 451–4.
1965. On the acoustic basis of the perception of intonation by linguists. *Word* 21: 40–54.
1967. *Intonation, perception, and language*. Cambridge, MA: MIT Press.
Lieberman, Philip and Sheldon B. Michaels. 1962. Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech. *JASA* 34: 922–7. Reprinted in Bolinger 1972a: pp. 235–49.
Lindau, Mona. 1986. Testing a model of intonation in a tone language. *JASA* 80: 757–64.
Lindsey, Geoffrey. 1985. *Intonation and interrogation: tonal structure and the expression of a pragmatic function in English and other languages*. PhD thesis, University of California, Los Angeles.
Lindström, Eva and Bert Remijsen. 2005. Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics* 43: 839–70.
McCawley, James D. 1978. What is a tone language? In Fromkin 1978: pp. 113–32.
McClure, J. Derrick. 1980. Western Scottish intonation: a preliminary study. In L. Waugh and C. van Schooneveld (eds.), *The melody of language*. Baltimore: University Park Press, pp. 201–17.
McGregor, Jeannette. 2005. *High rising tunes in Australian English*. PhD thesis, Macquarie University.
McNeill, David. 1992. *Hand and mind*. University of Chicago Press.
Maeda, Shinji. 1976. *A characterization of American English intonation*. PhD thesis, MIT.

Makarova, Veronika. 2007. The effect of pitch peak alignment on sentence type identification in Russian. *LgSp* 50: 385–422.
Matthews, Peter H. 1966. The concept of rank in Neo-Firthian linguistics. *JL* 2: 101–9.
Mayo, Catherine, Matthew Aylett, and D. Robert Ladd. 1997. Prosodic transcription of Glasgow English: an evaluation study of GlaToBI. In Botinis *et al.* 1997: pp. 231–4.
Mencken, H.L. 1948. American street names. *American Speech* 23: 81–8.
Menn, Lise and Suzanne Boyce. 1982. Fundamental frequency and discourse structure. *LgSp* 25: 341–83.
Mertens, Piet. 2002. *The prosogram: transcription of prosody using pitch contour stylization based on a tonal perception model and automatic segmentation*. Online at http://bach.arts.kuleuven.be/pmertens/prosogram (last accessed 22 March 2008).
Möbius, Bernd 1993. Perceptual evaluation of rule-generated intonation contours for German interrogatives. *Working Papers* 41 (Department of Linguistics and Phonetics, University of Lund): *Proceedings of the ESCA Workshop on Prosody, Lund, 27–29 September 1993*, 216–19.
Möbius, B., M. Pätzold, and W. Hess. 1993. Analysis and synthesis of German F_0 contours by means of Fujisaki's model. *SpComm* 13: 53–61.
Monaghan, A. I. C. 1991. *Intonation in a text to speech conversion system*. PhD thesis, University of Edinburgh.
1992. Heuristic strategies for the higher-level analysis of unrestricted text. In G. Bailly, G. Benoit, and T.R. Sawallis (eds.), *Talking machines: theories, models, and designs*. Amsterdam: Elsevier, pp. 143–61.
Mücke, Doris, Martine Grice, Johannes Becker, and Anne Hermes. Forthcoming. Sources of variation in tonal alignment: evidence from acoustic and kinematic data. Submitted to *JPhon*.
Nespor, Marina and Irene Vogel. 1982. Prosodic domains of external sandhi rules. In Harry van der Hulst and Norval Smith (eds.), *The structure of phonological representations*, vol. I. Dordrecht: Foris, pp. 225–55.
1983. Prosodic structure above the word. In Cutler and Ladd 1983: pp. 123–40.
1986. *Prosodic phonology*. Dordrecht: Foris. Republished 2007 by Mouton de Gruyter, Berlin.
Newman, Stanley. 1946. On the stress system of English. *Word* 2: 171–87.
Newmeyer, Frederick J. 2007. Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11: 133–57.
Nibert, Holly. 2000. *Phonetic and phonological evidence for intermediate phrasing in Spanish*. PhD thesis, University of Illinois, Urbana-Champaign.
Niebuhr, Oliver. 2007. The signalling of German rising-falling intonation categories – the interplay of synchronization, shape, and height. *Phonetica* 64: 174–93.

Nolan, Francis and Esther Grabe. 1997. Can ToBI transcribe intonational variation in English? In Botinis *et al.* 1997: pp. 259–62.
Nooteboom, S.G. and J.G. Kruyt. 1987. Accents, focus distribution and the perceived distribution of given and new information: an experiment. *JASA* 82: 1512–24.
Nooteboom, S.G. and J. M. B. Terken. 1982. What makes speakers omit pitch accents? An experiment. *Phonetica* 39: 317–36.
O'Connor, J.D. and G.F. Arnold. 1973. *Intonation of colloquial English*, second edition. London: Longman.
Odden, David. 1995. Tone: African languages. In J. Goldsmith (ed.), *The handbook of phonological theory*. Oxford: Blackwell, pp. 444–75.
Odé, Cecilia. 1989. *Russian intonation: a perceptual description*. Amsterdam: Rodopi.
1994. On the perception of prominence in Indonesian. In Odé and van Heuven 1994: pp. 27–107.
Odé, Cecilia and Vincent J. van Heuven (eds.), 1994. *Experimental studies of Indonesian prosody*. Department of Languages and Cultures of Southeast Asia and Oceania, University of Leiden.
Ofuka, Etsuko, J. Denis McKeown, Mitch Waterman, and Peter Roach. 2000. Prosodic cues for rated politeness in Japanese speech. *SpComm* 32: 199–217.
Ohala, John J. 1975. Review of Lehiste 1970. *Lg* 51: 736–40.
Ohala, John J. 1984. An ethological perspective on common cross-language utilization of F_0 in voice. *Phonetica* 41: 1–16.
Ohala, John J. 1994. The frequency code underlies the sound symbolic use of voice pitch. In L. Hinton, J. Nichols, and J. J. Ohala (eds.), *Sound symbolism*. Cambridge University Press, pp. 325–47.
Öhman, Sven. 1967. Word and sentence intonation: a quantitative model. *Speech Transmission Laboratory (Stockholm) Quarterly Progress and Status Report* 2: 20–54.
Ortiz-Lira, Hector. 1994. *A contrastive analysis of English and Spanish sentence accentuation*. PhD thesis, University of Manchester.
O'Shaughnessy, Douglas and Jonathan Allen. 1983. Linguistic modality effects on fundamental frequency in speech. *JASA* 74: 1155–71.
Palmer, Harold. 1922. *English intonation, with systematic exercises*. Cambridge: Heffer.
Partee, Barbara H. 1999. Focus, quantification, and semantics-pragmatics issues. In Bosch and van der Sandt 1999: pp. 213–31.
Patel, Aniruddh. 2007. *Music, language and the brain*. Oxford University Press.
Patterson, David. 2000. *A linguistic approach to pitch range modelling*. PhD thesis, University of Edinburgh.
Patterson, David and D. Robert Ladd. 1999. Pitch range modelling: Linguistic dimensions of variation. *ICPhS 14* (San Francisco): 1169–73.

Perlmutter, David. 1978. Impersonal passives and the unaccusative hypothesis. *BLS* 4: 157–89.
Pierrehumbert, Janet. 1980. *The phonology and phonetics of English intonation*. PhD thesis, MIT. Published 1988 by IULC, Bloomington, IN.
1981. Synthesizing intonation. *JASA* 70: 985–95.
1990. Phonological and phonetic representation. *JPhon* 18: 375–94.
2000. Tonal elements and their alignment. In Horne 2000: pp. 11–36.
Pierrehumbert, Janet and Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
Pierrehumbert, Janet, Mary E. Beckman, and D. Robert Ladd. 2000. Conceptual foundations of phonology as a laboratory science. In N. Burton-Roberts, P. Carr, and G.J. Docherty (eds.), *Conceptual and empirical foundations of phonology*. Oxford University Press, pp. 273–303.
Pierrehumbert, Janet and Julia Hirschberg. 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, pp. 271–311.
Pierrehumbert, Janet and Mark Liberman. 1982. Modelling the fundamental frequency of the voice. (Review of Cooper and Sorensen 1981.) *Contemporary Psychology* 27: 690–2.
Pierrehumbert, Janet and Shirley Steele. 1989. Categories of tonal alignment in English. *Phonetica* 46: 181–96.
Pike, Kenneth L. 1945. *The intonation of American English*. Ann Arbor: University of Michigan Press.
1948. *Tone languages: a technique for determining the number and type of pitch contrasts in a language, with studies in tonemic substitution and fusion*. Ann Arbor: University of Michigan Press.
Pitrelli, John F., Mary E. Beckman, and Julia Hirschberg. 1994. Evaluation of prosodic transcription labeling reliability in the ToBI framework. *ICSLP* 1 (Yokohama): 123–6.
Poser, William J. 1984. *The phonetics and phonology of tone and intonation in Japanese*. PhD thesis, MIT.
Post, Brechtje. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
Prevost, Scott and Mark Steedman. 1994. Specifying intonation from context for speech synthesis. *SpComm* 15: 139–53.
Price, Patti J., M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong. 1991. The use of prosody in syntactic disambiguation. *JASA* 90: 2956–70.
Prieto, Pilar. 1998. The scaling of the L values in Spanish downstepping contours. *JPhon* 26: 261–82.
2005. Stability effects in tonal clash contexts in Catalan. *JPhon* 33: 215–42.
Prieto, Pilar, Mariapaola D'Imperio, and Barbara Gili Fivela. 2005. Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *LgSp* 48: 359–96.

Prieto, Pilar and Francisco Torreira. 2007. The segmental anchoring hypothesis revisited. Syllable structure and speech rate effects on peak timing in Spanish. *JPhon* 35: 473–500.
Prieto, Pilar, Jan van Santen, and Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *JPhon*. 23: 429–51.
Prince, Alan. 1983. Relating to the grid. *LI* 14: 19–100.
Prince, Alan and Paul Smolensky. 2004. *Optimality theory: constraint interaction in generative grammar*. Oxford: Blackwell.
Prince, Ellen. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*. New York: Academic Press, pp. 223–56.
Reinholt Petersen, Niels. 1986. Perceptual compensation for segmentally conditioned fundamental frequency perturbation. *Phonetica* 43: 21–42.
Remijsen, Bert. 2002. Lexically contrastive stress accent and lexical tone in Ma'ya. In C. Gussenhoven and N. Warner (eds.), *Laboratory phonology VII*. Berlin/New York: Mouton de Gruyter, pp. 585–614.
Remijsen, Bert and Vincent van Heuven. 2005. Stress, tone, and discourse prominence in the Curaçao dialect of Papiamentu. *Phonology* 22: 205–35.
Repp, Bruno H. 1984. Categorical perception: issues, methods, findings. In N. Lass (ed.), *Speech and language*, vol. 10: *Advances in basic research and practice*. Orlando: Academic Press, pp. 244–335.
Rialland, Annie. 2007. Question prosody: an African perspective. In C. Gussenhoven and T. Riad (eds.), *Tones and tunes*, vol. I: *Typological studies in word and sentence prosody*. Berlin: Mouton de Gruyter, pp. 35–62.
Rialland, Annie and Stéphane Robert. 2001. The intonational system of Wolof. *Linguistics* 39: 893–939.
Rietveld, A.C.M. and Carlos Gussenhoven. 1985. On the relation between pitch excursion size and prominence. *JPhon* 13: 299–308.
1995. Aligning pitch targets in speech synthesis: effects of syllable structure. *JPhon* 23: 375–85.
Roach, Peter. 1994. Conversion between prosodic transcription systems: Standard British and ToBI. *SpComm* 15: 91–9.
Rooth, Mats. 1985. *Association with focus*. PhD thesis, University of Massachusetts.
1992. A theory of focus interpretation. *Natural Language Semantics* 1: 75–116.
Rose, Phil. 1987. Considerations in the normalisation of the fundamental frequency of linguistic tone. *SpComm* 6: 343–51.
1989. Phonetics and phonology of Yang tone phonation types in Zhenhai. *Cahiers de Linguistique Asie Orientale* 18: 229–45.
1990. Acoustics and phonology of complex tone sandhi. *Phonetica* 47: 1–35.
Rowlands, E.C. 1969. *Teach yourself Yoruba*. Sevenoaks, Kent: Hodder and Stoughton.

Rump, H. H. and René Collier. 1996. Focus conditions and the prominence of pitch-accented syllables. *LgSp* 39: 1–15.
Sag, Ivan and Mark Liberman. 1975. The intonational disambiguation of indirect speech acts. *CLS* 11: 487–97.
Schepman, Astrid, Robin Lickley, and D. Robert Ladd. 2006. Effects of vowel length and 'right context' on the alignment of Dutch nuclear accents. *JPhon* 34: 1–28.
Scherer, Klaus R., D. Robert Ladd, and K. Silverman. 1984. Vocal cues to speaker affect: testing two models. *JASA* 76: 1346–56.
Schiller, Niels F., Antje Meyer, and Willem Levelt. 1997. The syllabic structure of spoken words: evidence from the syllabification of intervocalic consonants. *LgSp* 40: 103–40.
Schmerling, Susan F. 1976. *Aspects of English sentence stress*. Austin: University of Texas Press.
Schubiger, Maria. 1958. *English intonation: its form and function*. Tübingen: Niemeyer.
1965. English intonation and German modal particles: a comparative study. *Phonetica* 12: 65–84. Reprinted in Bolinger 1972a: pp. 175–93.
1980. English intonation and German modal particles, II: a comparative study. In L. R. Waugh and C. H. van Schooneveld (eds.), *The melody of language*. Baltimore: University Park Press, pp. 279–98.
Scott, Donia R. 1982. Duration as a cue to the perception of phrase boundary. *JASA* 71: 996–1007.
Selkirk, Elisabeth O. 1980. The role of prosodic categories in English word stress. *LI* 11: 563–605.
1984. *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.
1986. On derived domains in sentence phonology. *Phonology Yearbook* 3: 371–405.
1995. Sentence prosody: intonation, stress, and phrasing. In J. Goldsmith (ed.), *The handbook of phonological theory*. Oxford: Blackwell, pp. 550–69.
Sharp, Alan E. 1958. Falling–rising intonation patterns in English. *Phonetica* 2: 127–52.
Shattuck-Hufnagel, Stefanie. 1995. The importance of phonological transcription in empirical approaches to 'stress shift' vs. 'early accent': comments on Grabe and Warren, and Vogel, Bunnell and Hoskins. In Connell and Arvaniti 1995: pp. 128–40.
Shattuck-Hufnagel, Stefanie, Mari Ostendorf, and K. Ross. 1994. Stress shift and early pitch accent placement in lexical items in American English. *JPhon* 22: 357–88.
Shattuck-Hufnagel, Stefanie and Alice E. Turk. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25: 193–247.

Shen, Xiao-nan Susan. 1990. *The prosody of Mandarin Chinese*. University of California Publications in Linguistics 118. Berkeley: University of California Press.
Shields, L. W. and D. A. Balota. 1991. Repetition and associative context effects in speech production. *LgSp* 34: 47–55.
Shriberg, Elizabeth, D. Robert Ladd, Jacques Terken, and Andreas Stolcke. 1996. Modeling pitch range variation within and across speakers: predicting F_0 targets when 'speaking up'. *ICSLP* 4 (Philadelphia): 650–3. Online version at www-speech.sri.com/cgi-bin/run-distill?papers/icslp96-scaling.ps.gz (last accessed 22 March 2008) replaces version printed in proceedings.
Silverman, Kim. 1986. F_0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica* 43: 76–91.
1987. *The structure and processing of fundamental frequency contours*. PhD thesis, University of Cambridge.
1990: The separation of prosodies: comments on Kohler's paper. In Kingston and Beckman 1990: pp. 139–51.
Silverman, Kim, Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg. 1992. ToBI: a standard for labeling English prosody. *ICSLP 2* (Banff): 867–70.
Silverman, Kim and Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In Kingston and Beckman 1990: pp. 71–106.
Slater, Andrew. 2007. *The IViE corpus: English Intonation in the British Isles*. Online at www.phon.ox.ac.uk/IViE/index.php (last accessed 22 March 2008).
Sledd, James. 1955. Review of Trager and Smith 1951. *Lg* 31: 312–35.
Sluijter, Agaath and Vincent van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *JASA* 100: 2471–85.
Smith, Henry Lee Jr. 1955. Review of W. Jassem, *Intonation of conversational English*. *Lg* 31: 189–93.
Sorace, Antonella. Forthcoming. *Gradience in split intransitivity*. Oxford University Press.
Sosa, Juan Manuel. 1999. *La entonación del español: su estructura fónica, variabilidad y dialectología*. Madrid: Ediciones Cátedra.
Steedman, Mark. 1991. Structure and intonation. *Lg* 67: 260–96.
2000. Information structure and the syntax-phonology interface. *LI* 31: 649–89.
Steele, Shirley. 1986. Nuclear accent F_0 peak location: effects of rate, vowel, and number of following syllables. *JASA* 80: S51.
Sugito, Miyoko and Hajime Hirose. 1988. Production and perception of accented devoiced vowels in Japanese. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics* 22: 21–39.
Swerts, Marc, Emiel Krahmer, and Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *JPhon* 30: 629–54.

Syrdal, Ann and Julia McGory. 2000. Inter-transcriber reliability of ToBI prosodic labelling. *ICSLP 6* (Beijing).
Szendrői, Kriszta. 2003. A stress-based approach to the syntax of Hungarian focus. *Linguistic Review* 20: 37–78.
Tan, N., Aiello, R., and Bever, T. G. 1981. Harmonic structure as a determinant of melodic organization. *Memory and Cognition* 9: 533–9.
Tartter, V. 1980. Happy talk: perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics* 27: 24–7.
Taylor, Paul A. 1995. The rise/fall/connection model of intonation. *SpComm* 15: 169–86.
2000. Analysis and synthesis of intonation using the Tilt model. *JASA* 107: 1697–714.
Terken, Jacques. 1991. Fundamental frequency and perceived prominence of accented syllables. *JASA* 89: 1768–76.
Terken, Jacques and Julia Hirschberg. 1994. Deaccentuation of words representing given information: effects of persistence of grammatical function and surface position. *LgSp* 37: 125–45.
't Hart, Johan. 1979. Explorations in automatic stylization of F_0 curves. *IPO Annual Progress Report* 14: 61–5.
1981. Differential sensitivity to pitch distance, particularly in speech. *JASA* 69: 811–21.
't Hart, Johan and Antonie Cohen. 1973. Intonation by rule: a perceptual quest. *JPhon* 1: 309–27.
't Hart, Johan and René Collier. 1975. Integrating different levels of intonation analysis. *JPhon* 3: 235–55.
't Hart, Johan, René Collier, and Antonie Cohen. 1990. *A perceptual study of intonation: an experimental-phonetic approach*. Cambridge University Press.
Thorsen, Nina. *See also* Grønnum.
1980. A study of the perception of sentence intonation: evidence from Danish. *JASA* 67: 1014–30.
1983. Two issues in the prosody of Standard Danish. In Cutler and Ladd 1983: pp. 27–38.
1985. Intonation and text in Standard Danish. *JASA* 77: 1205–16.
1986. Sentence intonation in textual context: supplementary data. *JASA* 80: 1041–7.
Trager, George L. 1958. Paralanguage: a first approximation. *Studies in Linguistics Occasional Papers* 13: 1–12.
1964. The intonation system of American English. In David Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. M. Trim (eds.), *In honour of Daniel Jones*. London: Longman. Reprinted in Bolinger 1972a: pp. 83–6.
Trager, George L. and H. L. Smith. 1951. *An outline of English structure*. Norman, OK: Battenburg Press. Reprinted 1957 by American Council of Learned Societies, Washington.

Traunmüller, H. and A. Eriksson. 1994. The perceptual evaluation of F_0 excursions in speech as evidenced in liveliness estimations. *JASA* 97: 1905–15.
Trim, J. L. M. 1959. Major and minor tone-groups in English. *Le Maître Phonétique* 112: 26–9.
Truckenbrodt, Hubert. 1999. On the relation between syntactic phrases and phonological phrases. *LI* 30: 219–55.
2002. Upstep and embedded register levels. *Phonology* 19: 77–120.
2004. Final lowering in non-final position. *JPhon* 32: 313–48.
Uhmann, Susanne. 1988. Akzenttöne, Grenztöne und Fokussilben: zum Aufbau eines phonologischen Intonationssystems für das Deutsche. In Altmann 1988: pp. 65–88.
1991. *Fokusphonologie: eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Tübingen: Niemeyer.
Uldall, Elizabeth. 1964. Dimensions of meaning in intonation. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. M. Trim (eds.), *In honour of Daniel Jones*. London: Longman. Reprinted in Bolinger 1972a: pp. 250–9.
Underhill, Robert. 1976. *Turkish grammar*. Cambridge, MA: MIT Press.
Vaissière, Jacqueline. 1983. Language-independent prosodic features. In Cutler and Ladd 1983: pp. 53–66.
Vallduví, Enric. 1990. *The informational component*. PhD thesis, University of Pennsylvania. Published 1992 by Garland Press, New York.
1991. The role of plasticity in the association of focus and prominence. *Proceedings of the Eastern States Conference on Linguistics* (ESCOL) 7: 295–306.
Vallduví, Enric and Maria Vilkuna. 1998. On rheme and kontrast. In P. Culicover and L. McNally (eds.), *Syntax and semantics, vol. XXIX: The limits of syntax*. San Diego: Academic Press, pp. 79–108.
Vallduví, Enric and Ron Zacharski. 1994. Accenting phenomena, association with focus, and the recursiveness of focus-ground. In P. Dekker and M. Stokhof (eds.), *Proceedings of the Ninth Amsterdam Colloquium* (Department of Philosophy, University of Amsterdam), pp. 683–702.
van den Berg, Rob, Carlos Gussenhoven, and Toni Rietveld. 1992. Downstep in Dutch: implications for a model. In Docherty and Ladd 1992: pp. 335–58.
Vanderslice, Ralph and Peter Ladefoged. 1972. Binary suprasegmental features and transformational word-accentuation rules. *Lg* 48: 819–38.
Vanderslice, Ralph and Laura S. Pierson. 1967. Prosodic features of Hawaiian English. *Quarterly Journal of Speech* 53: 156–66. Reprinted in Bolinger 1972a: pp. 439–50.
Varga, László. 1989. The stylized fall in Hungarian. *Acta Linguistica Hungarica* 39: 317–30.
2002. *Intonation and stress: evidence from Hungarian*. Basingstoke: Palgrave Macmillan.

Veilleux, Nanette, Alejna Brugos, and Stefanie Shattuck-Hufnagel. 2007. *The new English ToBI home page*. Online at www.tobihome.org (last accessed 22 March 2008).
Venditti, Jennifer, Sun-Ah Jun, and Mary Beckman. 1996. Prosodic cues to syntactic and other linguistic structures in Japanese, Korean and English. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax: bootstrapping from speech to grammar in early acquisition*, Part IV: *Speech and the acquisition of phrase structure*. Hillsdale, NJ: Lawrence Erlbaum, pp. 287–311.
Verhoeven, Jo. 1994. The discrimination of pitch movement alignment in Dutch. *JPhon* 22: 65–85.
Vogel, Irene and Istvan Kenesei. 1987. The interface between phonology and other components of grammar: the case of Hungarian. *Phonology Yearbook* 4: 243–63.
von Stechow, Arnim and Susanne Uhmann. 1986. Some remarks on focus projection. In W. Abraham and S. de Meij (eds.), *Topic, focus, and configurationality*. Amsterdam: Benjamins, pp. 295–320.
Wagner, Michael. 2005. *Prosody and recursion*. PhD thesis, MIT.
Wang, William. 1967. The phonological features of tone. *IJAL* 33: 93–105.
Ward, Gregory and Julia Hirschberg. 1985. Implicating uncertainty. *Lg* 61: 747–76.
Ward, Ida C. 1933. *The phonetic and tonal structure of Efik*. Cambridge: Heffer.
Warren, Paul. 2005. Patterns of late rising in New Zealand English: intonational variation or intonational change? *LVC* 17: 209–30.
Warren, Paul and David Britain. 2000. Intonation and prosody in New Zealand English. In Allan Bell and Koenraad Kuiper (eds.), *New Zealand English*. Wellington: Victoria University Press, pp. 146–72.
Wedgwood, Daniel. 2005. *Shifting the focus: from static structures to the dynamics of interpretation*. Oxford: Elsevier.
Weinreich, Uriel. 1954. Stress and word structure in Yiddish. In Uriel Weinreich (ed.), *The field of Yiddish: studies in Yiddish language, folklore and literature*. Linguistic Circle of New York, pp. 1–27.
Welby, Pauline. 2003. Effects of pitch accent position, type, and status on focus projection. *LgSp* 46: 53–81.
Wells, J. C. 1982. *Accents of English*, vol. I: *An introduction*; vol. II: *The British Isles*. Cambridge University Press.
Wells, Rulon. 1945. The pitch phonemes of English. *Lg* 21: 27–40.
Welmers, William. 1973. *African language structures*. Berkeley: University of California Press.
Whalen, D. H., Arthur S. Abramson, Leigh Lisker, and M. Mody. 1990. Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica* 47: 36–49.
Whalen, D. H. and Andrea Levitt. 1995. The universality of intrinsic F_0 of vowels. *JPhon* 23: 349–66.

Wilbur, Ronnie 1994a. Eyeblinks and ASL phrase structure. *Sign Language Studies* 84: 221–40.
1994b. Foregrounding structures in ASL. *Journal of Pragmatics* 22: 647–72.
Willems, Nico, René Collier, and Jan Roelof de Pijper. 1988. A synthesis scheme for British English intonation. *JASA* 84: 1250–61.
Williams, Carl E. and Kenneth N. Stevens. 1972. Emotions and speech: some acoustical correlates. *JASA* 52: 1238–50.
Williams, Edwin. 1976. Underlying tone in Margi and Igbo. *LI* 7: 463–84.
Winston, F. D. D. 1960. The 'mid' tone in Efik. *African Language Studies* 1: 185–92.
Woo, Nancy. 1969. *Prosody and phonology*. PhD thesis, MIT. Distributed by IULC.
Wunderlich, Dieter. 1988. Der Ton macht die Melodie: zur Phonologie der Intonation des Deutschen. In Altmann 1988: pp. 1–40.
Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.
2005. Speech melody as articulatorily implemented communicative functions. *SpComm* 46: 220–51.
Xu, Yi and Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *JASA* 111: 1399–413.
Xu, Yi and Q. Emily Wang. 2001. Pitch targets and their realization: evidence from Mandarin Chinese. *SpComm* 33: 319–37.
Xu, Yi and Ching X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *JPhon* 33: 159–97.
Yip, Moira. 1989. Contour tones. *Phonology* 6: 149–74.
2002. *Tone*. Cambridge University Press.
Yoon, Tae-Jin. 2007. *A predictive model of prosody through grammatical interface: a computational approach*. PhD thesis, University of Illinois, Urbana-Champaign.
Yuen, Ivan. 2003. *Downtrend and normalisation*. PhD thesis, University of Edinburgh.
Zsiga, Elizabeth and Rattima Nitisaroj. 2007. Tone features, tone perception, and peak alignment in Thai. *LgSp* 50: 343–83.
Zubizarreta, Maria Luisa. 1998. *Prosody, focus, and word order*. Cambridge, MA: MIT Press.

Index of names

As a rule, names are not included in the following list if the only reference to them is as second or subsequent author, as editor of a collection of papers or website, or in passing mentions in footnotes and figure captions.

Index of languages

Subject index