Educational Linguistics

Jesús Romero-Trillo *Editor*

Pragmatics and Prosody in English Language Teaching

Springer

Pragmatics and Prosody in English Language Teaching

Educational Linguistics

Volume 15

The *Educational Linguistics* book series focuses on work that is: innovative, trans-disciplinary, contextualized and critical.

In our compartmentalized world of diverse academic fields and disciplines there is a constant tendency to specialize more and more. In academic institutions, at conferences, in journals, and in publications the crossing of disciplinary boundaries is often discouraged.

This series is based on the idea that there is a need for studies that break barriers. It is dedicated to innovative studies of language use and language learning in educational settings worldwide. It provides a forum for work that crosses traditional boundaries between theory and practice, between micro and macro, and between native, second and foreign language education. The series also promotes critical work that aims to challenge current practices and offers practical, substantive improvements.

For further volumes:
http://www.springer.com/series/5894

Jesús Romero-Trillo
Editor

Pragmatics and Prosody in English Language Teaching

Springer

Editor
Jesús Romero-Trillo
Facultad de Filosofía y Letras
Universidad Autónoma de Madrid
C/Francisco Tomás y Valiente 1
Madrid, Madrid
Spain

ISSN 1572-0292
ISBN 978-94-007-3882-9 e-ISBN 978-94-007-3883-6
DOI 10.1007/978-94-007-3883-6
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2012933584

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Amanda Baker is a Lecturer in TESOL in the Faculty of Education at the University of Wollongong, Australia. Her research interests focus on second language learning and teaching, including prosodic analyses of L2 speech, pronunciation pedagogy, oral communication pedagogy, classroom research and teacher education.

Heather L. Balog is an Assistant Professor in the Department of Communication Sciences and Disorders at Wayne State University in Detroit, MI. Her research focuses on the intonation production characteristics in young children and how these characteristics relate to pragmatic aspects of language development such as communicative intentions.

David Deterding is an Associate Professor in the Faculty of Arts and Social Sciences at the University of Brunei Darussalam, where he teaches phonetics, grammar, Chinese-English translation, research methods and introductory linguistics. His research has focused on the acoustic measurement of speech, the pronunciation of Malay, and a description of English in East Asia, especially the varieties spoken in Singapore, China, Hong Kong and Brunei. His most recent book is *Singapore English*, published by Edinburgh University Press in 2007.

Francisco Gutiérrez Díez obtained his Ph.D. in English Philology from the University of Barcelona and is Professor of English Philology at the Faculty of Arts (University of Murcia). His research interests are the application of comparative phonetics to language teaching/learning). He is an expert member of the Language Policy Committee of the World Animal Health Organization (OIE) and was the linguistic coordinator of the translation into Spanish of the Manual of diagnostic tests and vaccines for terrestrial animals (OIE, Paris, 2004), and the Manual of diagnostic tests for aquatic animals (OIE, Paris, 2006).

Guiling (Gloria) Hu received her Ph.D. from the Department of Applied Linguistics & ESL at the Georgia State University in 2009. Her research interests include second language acquisition, cross-linguistic study of prosody, sociolinguistics, and psycholinguistics. She is currently working at the National Foreign Language Center of the University of Maryland.

Angelos Lengeris, Ph.D. University College London, is a postdoctoral researcher at the Aristotle University of Thessaloniki, Greece. His research interests include speech perception and production, acoustic phonetics, the acquisition of L2 intonation, and computer-assisted training for L2 learners.

Jessica Newell is currently an instructor at Johns Hopkins University Center for Talented Youth teaching Cognitive Psychology. She has studied language, communication, attraction, and motivation and has worked in several different research laboratories (both in Spain and the USA). She has an expressed interest in communication and what motivates individuals. She earned her B.A (Psychology) and M.A. (Experimental Research) at Cleveland State University graduating summa cum laude in 2011.

Marie Nilsenová is an assistant professor in the Department of Communication and Information Sciences at the Tilburg University. Her past research focused on the semantics of prosody, in particular the use of intonation as a cue to various social traits. In her current work, she is exploring individual differences in the perception of pitch and their effect on phoneme categorization and emotion identification.

Lucy Pickering is an Associate Professor in the Department of Literature and Languages at Texas A&M-Commerce and is the director of the Applied Linguistics Laboratory. She received her Ph.D. from the University of Florida in 1999. Her research interests include the applications of speech analysis to discourse analysis, the cross-linguistic transfer of prosodic features and the teaching of intonation in EFL.

Thorsten Piske, Ph.D. Kiel University, is professor and chair of Teaching English as a Foreign Language at the Friedrich-Alexander-University of Erlangen-Nuremberg. His research focuses on first and second language acquisition and on bilingual education. He has published studies on the production and perception of first and second language speech sounds and on the effectiveness of bilingual programs in preschools and elementary schools.

Beatrice Szczepek Reed is a Lecturer in Language Education in the Department of Education at the University of York, UK. She has published widely in the areas of Conversation Analysis, prosody in interaction, and English speaking skills and pronunciation. She is also the author of the textbook *Analysing Conversation: An Introduction to Prosody* (Palgrave, 2010).

Silvia Riesco-Bernier holds a permanent position at the Official Language School in Madrid as an English teacher and EFL teacher-trainer. She gained a Ph.D. in Applied Linguistics in 2007 and has been teaching phonetics and phonology at Universidad Autónoma de Madrid. She is the coordinator of CLAN project and is a member of the editorial board of TESOL-Spain, Classroom Discourse and *Journal of Multicultural Discourses*.

Jesús Romero-Trillo is an Associate Professor in the Department of English Philology at Universidad Autónoma de Madrid. He specializes in the pragmatics-intonation interface in English conversations. The Editor-in-Chief of the 'Yearbook of Corpus

Linguistics and Pragmatics' (Springer). He is also a member of the editorial board of *Journal of Pragmatics* and has been the Review Editor of the journal *Intercultural Pragmatics* since its inception. He has edited the volume *Pragmatics and Corpus Linguistics: A Mutualistic Entente*, (Mouton de Gruyter 2008), and is the director of the CLAN (Corpus of Language and Nature) Project.

Marc Swerts is a full professor in the Department of Communication and Information Sciences at the Tilburg University. He is specialized in research on communicative functions of audiovisual prosody in speech and explores the perception of audiovisual cues in different age groups and cultures, both in human-human and human-computer interactions, and their utilization in first and second language acquisition. In recent projects, he has been focusing on the role of prosody in autism and as a potential cue to depression.

Tim Wharton is based at Kingston University. His research focuses on 'natural' communicative behaviours and his main theses outlined in a 2009 book. Prior to his academic career, he was a singer-songwriter and has written and recorded a number of songs still used as teaching resources around the world for those learning English as a foreign language.

Yi Xu, Reader in Speech Science at University College London, UK. Ph. D. in Linguistics from the University of Connecticut, USA, Postdoc at MIT and Assistant Professor at Northwestern University. He has published widely since 1986, covering topics on the production, perception and theoretical modeling of tone, intonation, segment and the syllable.

Chapter 1
Introduction

Jesús Romero-Trillo

The volume 'Pragmatics and Prosody in English Language Teaching' is an attempt to bridge the (often) existing gap between the study of pragmatic meaning and the study of prosodic features in real interaction. Pragmatics, as a relatively new but at the same time multifaceted discipline of linguistics, covers numerous aspects of theoretical and applied approaches to real language and its multiple interpretations. Its original tenets have related to the interpretation of discourse in specific contexts, and in this sense, it has relied on detailed descriptions of context to disambiguate possible meanings. In this sense, context was studied as the crucial element for meaning but often in isolation and most research aimed at the justification or exemplification of theoretical approaches. In recent times, however, some linguists have advocated the use of contextualized language data to study pragmatics theoretically (cf. Romero-Trillo 2008; Jucker et al. 2009) and practically (O'Keefe et al. 2011).

This volume proposes another turn of the screw and defends the absolute need to incorporate prosody into the pragmatic analysis of speech. In other words, the foundations of this volume lie in the belief that the study of context and meaning is incomplete without the careful analysis of the acoustic elements that compose the kaleidoscope of speech. In some way, this approach is a return to the origins of discourse analysis and pragmatics in the late 1960s and early 1970s, when revered functional grammarians and discourse analysts like Halliday (1967, 1970) and Brazil (1975), *inter alios*, demonstrated the inextricable relationship between prosody and language functions. Some years later, Levinson highlighted this liaison as a means to 'tame' the power of grammar in real interaction: 'Various syntactic rules seem to be properly constrained only if one refers to pragmatic conditions; and similarly for matters of stress and intonation' (1983: 36).

J. Romero-Trillo (✉)
Departamento de Filología Inglesa, Facultad de Filosofía,
Universidad Autónoma de Madrid, Madrid, Spain
e-mail: jesus.romero@uam.es

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_1,

The way an utterance is pronounced belongs to the realm of prosody and the acoustic analysis of prosody can be very complex, as it requires advanced knowledge of acoustic theory to assess individual differences between speakers in terms of sex, age, emotional state, dialectal origin, etc. The basic acoustic parameters that measure these individual features are the spectrum, the duration, the intensity, the formants, the pulses and the pitch level. All these elements portray specific features that allow linguists to describe utterances in great detail. Based on the acoustic performance of language, pragmatics intends to identify the intentions with which utterances are pronounced and how they may help clarify the meaning behind some grammatical structures that do not render their transparent pragmatic force on the basis of their construction.

Although nowadays it is possible to obtain a good description of prosody with the aid of computer programmes such as Praat (Boersma and Weenink 2010), pragmaticians have rarely approached the study of meaning from a prosodic perspective. Sometimes, it seems as if speech were the subject matter for the study of meaning, as in a script, but without the study of that script in a context. The combined study of prosody and pragmatics necessarily demands the use of acoustic analysis to identify the elements that are significant for meaning creation at the pragmatics level. With these tools, pragmatics needs to differentiate between the useful features in the description of the individual speaker and those that knit the web of meaning contrasts at the language level.

This combination of pragmatics and prosody in language research is decisive for the analysis of real language. With the respectful distance of more than three decades of insightful and persistent analysis in language use, I believe that linguists nowadays face new demands for more delicate research on the acoustics of language to understand the ontology of pragmatics, especially in relation to the teaching of English to speakers of other languages. Recent approaches to pragmatics have emphasized the need for cross-cultural and intercultural aspects of pragmatics, with a clear emphasis on the different ways in which languages realize functions that are quasi-universal and how speakers of a second or foreign language can realize these functions in the target language. In my opinion, understanding this process is not a scholastic enterprise but an essential task for teaching a second or foreign language. In fact, equivocal realizations of certain functions can lead to pragmatic misunderstanding and the lack of this awareness in language teaching can lead to pragmatic fossilization (Romero-Trillo 2002).

The contributors to this volume are experts in prosody and pragmatics and have been working on the juxtaposition of both disciplines for many years. They approach the interface of pragmatics and prosody with a dynamic and prospective orientation based on a thorough revision of the literature in each field. Linguists, teacher-trainers and language teachers will find in this volume some of the most recent research developments in pragmatics and prosody and useful keys for comparison between native and non-native speech performance with orientations for ELT context.

The volume is organized in three parts: ‘Theoretical approaches to the teaching of Prosody’; ‘Pragmatics, Prosody and Communication’ and ‘Pedagogical implications for English Language Teaching’. As can be observed, the three parts represent a

cline from the most theoretically-oriented presentation of prosody to the most applied and classroom-oriented research, with some of the questions that ELT practitioners face when they try to explain the intonation, rhythm and stress of English to (non-) native speakers. The volume makes a necessary stop to show prosody and pragmatics as two essential sides of the same coin, with the proviso that there cannot be meaningful communication without their factual alliance.

The first part, Theoretical approaches to the teaching of prosody, departs with David Deterding and his chapter 'Issues in the acoustic measurement of rhythm'. The contribution describes one of the classic debates on the teaching of prosody: the division between stress-timed and syllable-timed languages. For the author, the claim that English belongs to the former variety of languages is not supported by recent research, which is more inclined to the existence of different types of rhythm in English according to the register and sociolinguistic and regional features of the different types of English spoken around the world.

The next chapter 'Prosody and second language teaching: Lessons from L2 speech perception and production research', by Angelos Lengeris, reviews historic findings from L2 speech perception and production research. For instance, he explains that L2 intonation learning is not restricted to childhood as is the popular belief, and shows how available freeware computer programmes can be used for teaching L2 intonation with outstanding results.

The chapter 'Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages', by Thorsten Piske, describes the concept of foreign accent and its negative effects not only on intelligibility but, especially, on the social acceptance of utterances produced by second language (L2) learners. The author classifies the degree of foreign accent into three aspects: segmental errors, suprasegmental errors and the lack of fluency due to pauses, hesitation phenomena and rate of speech. The chapter depicts the attitudes of native and non-native speakers towards foreign-accented speech and analyses the influence of the three elements mentioned above on the perception of foreign accent.

The chapter, 'Function vs. form in speech prosody—Lessons from experimental research and potential implications for teaching', by Yi Xu, departs from applications of the study of tone languages for the description of non-tone languages. According to the author, this liaison between different language models can help to understand the link between function and form in speech prosody, especially in the analysis of pitch, as both types of languages share a similar contour production. Yi Xu applies this notion to the comparison of lexical and extra-lexical functions in Mandarin and English. The chapter ends with useful advice on how prosody can be better approached in English language teaching.

The last chapter of the first part of the volume, 'Prosodic adaptation in language learning', by Marie Nilsenová and Marc Swerts, describes intonation, rhythm and accentuation as the acoustic basis for language acquisition. These prosodic elements are responsible for the exchange of information in discourse and also intervene in the structure of social regulations, e.g., to assign and accept group membership. The authors summarize current experimental findings in the area of prosodic adaptation

and show the link to first language acquisition and second language learning. Their contribution demonstrates that prosodic adaptation can contribute to social membership and language processing, to conclude that they are essential elements to promote in second language acquisition.

The second part of the volume, 'Pragmatics, Prosody and Communication', starts with the contribution by Tim Wharton, 'Prosody and meaning: Theory and practice', which approaches the function of stress and intonation in the creation of 'natural' and proper linguistic input. The chapter describes the challenges for pragmatics in terms of the characterization of prosody and in the relation with intentional communication. The author finishes by extending his theoretical tenets to the practical domain for the teaching of English pronunciation.

The next chapter, by Jesús Romero-Trillo and Jessica Newell, studies the realization of feedback by Pragmatic Markers as the 'go-ahead signals' that verify the correct reception of a message in an interaction. Their study compares the realization of native and non-native prosodic performance of feedback elements in a spoken corpus with statistical analyses. From a pedagogical perspective, their study of the acoustics of feedback in conversation is essential to understand how these elements function as 'punting poles', which sail through the flow of conversation and how foreign speakers of English need to master the prosody of these elements to be pragmatically correct.

The contribution by Heather Balog, 'Early prosodic production: Pragmatic and acoustic analyses for L2 language learners' discusses the co-occurrence of the development of prosodic speech characteristics *vis-a-vis* the development pragmatic language skills in young children. The chapter makes an overall review of the stages of pragmatic development in the early years and describes the process of intonation shaping towards the adult model. The author emphasizes the relevance of this awareness for second language development and makes suggestions for future research in this field.

The last chapter in this part, 'Prosody in conversation: Implications for teaching English pronunciation', by Beatrice Szczepek Reed, delves into the relationship between prosodic form and interactional function, especially with respect to prosody and turn taking and the role of prosody for interactional alignment. The author departs from the hypothesis that conversational cues work as clusters and that participants in conversation make prosodic choices in terms of social actions and not on the basis of abstract context-free functions of prosodic patterns.

The last part of the volume, 'Pedagogical implications for English language teaching', starts with the chapter 'Same but different: The pragmatic potential of native *vs*. non-native teachers' intonation in the EFL classroom' by Silvia Riesco-Bernier. The contribution investigates the pragmatics of intonation in teacher talk in a pre-school spoken EFL corpus. The chapter analyses the multifunctionality of prosody, evaluates the communicative functions displayed in the classroom and compares the prosodic choices made by native and non-native teachers. The findings show the correlation between communicative functions and prosodic realizations in the two groups of teachers, although through different intonation strategies.

Lucy Pickering, Guiling Hu and Amanda Baker's chapter, 'The pragmatic function of intonation: Cueing agreement and disagreement in spoken English discourse and implications for ELT', investigates the relationship between pitch and prosodic (mis) matching to indicate (dis) agreement in native speakers of American English and in Chinese learners of English. The contribution illustrates the cross-cultural manifestations of speech acts in relation to second language intonation acquisition.

The penultimate chapter 'Trouble spots in the learning of English intonation by Spanish speakers: Tonality and tonicity', by Francisco Gutiérrez Díez, focuses on some typical intonation errors by Spanish learners of English, with special attention to the intonation subsystems of tonality and tonicity (including errors of onset misplacement). The author offers some useful advice for the explicit pedagogical treatment of intonation errors in Spanish speakers of English and for the awareness of pragmatic meaning in relation to the presence or absence of pitch accent in speech segments.

The last chapter, written by Jesús Romero-Trillo and entitled 'Teaching prosody with a pragmatic orientation: A synthesis', approaches the topics discussed in the book at the pragmatics and prosody interface. Readers will find this overview very useful, as it highlights some aspects that need consideration in the English teaching context.

In conclusion, the present volume deals with the complex topic of the analysis of prosody and pragmatics in ELT practice. I believe that the necessary liaison between pragmatics and prosody, in theoretical and practical terms, can shed light on some of the difficulties that speakers of English as a second or foreign language face in daily communication. Students and teachers often concentrate on grammatical, lexical and even discourse aspects of English but rarely do they strive in the pronunciation of utterances with prosodic accuracy. This not only results in a lack of native-like pragmatic behaviour, with the subsequent risks of miscommunication, but it can also lead to what can be called 'performance insecurity' in their interaction with native speakers. I am certain that the chapters in this volume, with their careful synergy between current theoretical approaches to prosody and pragmatics, will help linguists and language teachers to tackle the often avoided, but approachable, challenge of teaching English prosody.

References

Boersma, Paul, and David Weenink. 2010. Praat: Doing phonetics by computer [Computer programme]. Version 5.1.07, from http://www.praat.org/.

Brazil, David. 1975. *Discourse intonation 1/2*. Birmingham: Birmingham University Press.

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

Halliday, M.A.K. 1970. *A course in spoken English*. Oxford: University Press.

Jucker, Andreas, Daniel Schereier, and Marianne Hundt, eds. 2009. *Corpora: Pragmatics and discourse*. Amsterdam: Rodopi.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

O'Keefe, Anne, Svenja Adolphs, and Brian Clancy. 2011. *Introducing pragmatics in use*. London: Routledge.
Romero-Trillo, Jesús. 2002. The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics* 34: 769–784.
Romero-Trillo, Jesús, ed. 2008. *Pragmatics and corpus linguistics, a mutualistic entente*. Berlin: Mouton de Gruyter.

Part I
Theoretical Approaches to the Teaching of Prosody

Chapter 2
Issues in the Acoustic Measurement of Rhythm

David Deterding

2.1 Introduction

In the past, it was often claimed that there are two fundamental kinds of rhythm in languages: stress-timed, where the rhythmic beat occurs on stressed syllables and the duration between successive stressed syllables tends to be approximately even; and syllable-timed, where the syllable constitutes the basis of the rhythmic beat and individual syllables tend to be more evenly timed. For example, Abercrombie (1967, p. 97) stated: "As far as is known, every language in the world is spoken with one kind of rhythm or with the other." While this is certainly an oversimplification, as it overlooks languages such as Japanese where the rhythm is based on a smaller unit than the syllable, the mora (Hoequist 1983), the belief in a contrast between the two basic kinds of rhythm is still widely held, especially among language teachers.

However, the strong claim about fixed rhythmic categories is generally no longer maintained among researchers and it is more common to regard the rhythm of languages as existing along a cline of stress/syllable timing (Miller 1984). Furthermore, we should observe that although various attempts have been made to classify languages into one category or another (e.g., Ramus et al. 1999), a single language can actually have various styles of pronunciation, so Crystal (1995) notes that although British English might usually be stress-timed, it also exhibits syllable-timing in some circumstances, such as baby talk, television commercials, some popular music, and expressing some emotions such as irritation and sarcasm.

In fact, nowadays many researchers no longer refer to 'timing' for rhythmic classification, as rhythm is not just about timing. An alternative terminology refers to 'stress-based' and 'syllable-based' rhythm, as the two categories crucially involve other things than just timing, including the tendency for an alternation between

D. Deterding (✉)
Faculty of Arts and Social Sciences, University of Brunei Darussalam,
Bandar Seri Begawan, Brunei Darussalam
e-mail: dhdeter@gmail.com

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_2,

strong and weak syllables or more broadly "the hierarchical organization of temporally coordinated prosodic units" (Cummins and Port 1998, p. 145). Indeed, one of the fundamental goals of metrical phonology is the prediction of the relative strength of each syllable (Hogg and McCully 1987). In this paper, reference will henceforth be made to stress-/syllable-based rhythm, even when discussing early work that actually used the timing terminology.

Although stress-based rhythm is typically found in most Inner-Circle varieties of English (using the circles model of Kachru 1985), syllable-based rhythm seems to be common in many Outer-Circle varieties, those Englishes that have emerged in post-colonial societies (Schneider 2007). For example, syllable-based rhythm has been reported in Singapore English (Brown 1988), India (Kachru 2006, p. 46), the Caribbean, and West Africa (Wells 1982, pp. 570, 639). Indeed, Crystal (2003, p. 171) speculates that syllable-based rhythm might one day become the norm for all varieties of English.

In drawing an analogy with recent developments in hip-hop culture around the world, Pennycook (2007) suggests that the greatest influence on the future of English no longer lies with the places where the language originated, such as Britain and America, but in other places where vibrant new forms and innovative patterns of usage are developing. And this observation is consistent with the suggestion that the syllable-based rhythm of Outer-Circle Englishes is likely to have an increasingly important influence over the future development of English worldwide, even if speakers in Inner-Circle societies may continue to feel uncomfortable about such developments.

Despite this widespread interest about the rhythm of English and other languages around the world, there remain fundamental issues about how to measure it acoustically. This paper will focus on the acoustic measurement of rhythm, using a slightly modified version of the metric developed by Low et al. (2000). It will consider the issues that arise when the metric is used and how they can be dealt with, and it will discuss the effectiveness of the metric in contrasting the rhythm of two varieties of English, those of Brunei and Britain, in particular by evaluating how reliable the measurements are and what they really show.

Finally, the paper will consider the teaching of rhythm; specifically whether it is appropriate to teach stress-based rhythm to students of English when syllable-based rhythm is so common in the Englishes spoken around the world.

2.2 The Acoustic Measurement of Rhythm

Early attempts to show an acoustic distinction between stress- and syllable-based languages were generally unsuccessful. For example, Roach (1982) investigated the timing distinctions between two groups of languages: French, Telugu and Yoruba, which are all claimed to be syllable-based; and English, Russian and Arabic, which are all archetypical stress-based languages. But he failed to find any difference between the timing of these two groups of languages.

However, more recently, acoustic measurements have demonstrated that rhythmic differences between languages or language varieties do exist. For example, Low et al. (2000) developed a metric called the Pairwise Variability Index (PVI), which is based on the comparison of successive vowel durations, and they showed that there is a difference in the rhythmic timing of sentences read by speakers of Singapore and British English; and Deterding (2001) used a metric that compared the duration of successive syllables in conversational speech and similarly showed that there is a significant difference between the rhythm of conversational Singapore English and British English.

The formula for the PVI suggested by Low et al. (2000) is as follows:

$$PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| \Big/ (m-1) \right]$$

where d is the duration of a vowel and m is the number of syllables in the utterance. In brief, the PVI obtains the difference between the duration of one vowel and the following vowel and then normalizes the result by dividing by their average duration (in order to ensure the result is not dependent on speaking rate). The final PVI value is derived by obtaining the average for all the syllables that are compared and multiplying by 100.

The PVI has been widely adopted in recent years in acoustic investigations into the rhythm of various languages, including French, Polish, Dutch, Thai and Malay (Grabe and Low 2002), Latvian (Bond et al. 2003), German and Spanish (Lleó et al. 2007), Czech (Dankovičová and Dellwo 2007), and Chinese (Benton et al. 2007). However, it is not without its problems: Barry and Russo (2003) report that the PVI gives a lower value for fast speaking rates, especially with varieties of Italian spoken in Naples and Pisa; Ong et al. (2005) show that the PVI results are highly sensitive to the decisions made by different measurers, as identifying the boundary between two syllables, such as in the middle of a sequence such as "more of", is almost entirely subjective, and they recommend that in any comparison of the rhythm of two varieties of speech, the two sets of measurements should be done by the same researcher to try to ensure that the same criteria are applied in both cases; and Nolan and Asu (2009) raise issues about how the PVI is used, particularly whether it is appropriate to depend entirely on comparison of vowel durations for the evaluation of rhythm, and they show that some languages, particularly Estonian, may be characterized by two co-existent kinds of rhythm, one based on syllables and the other on rhythmic feet, the period between two successive stressed syllables.

The current study compares the English spoken by undergraduates in Brunei with that of RP British speakers using a slightly modified version of the PVI that was proposed by Low et al. (2000). Three modifications to the PVI will be discussed in the next section. In addition, this paper will consider the reliability of the PVI results by comparing measurements made several months apart. Then issues arising in the application of the metric will be considered, before we conclude by discussing the teaching of rhythm.

2.3 Modifications to the PVI

Three modifications to the PVI are discussed here: treatment of very short syllables; omission of the final syllable of an utterance; and how to deal with approximants in consonant clusters.

The PVI depends on the measurement of vowel duration. However, in connected speech the vowel may be entirely omitted from some syllables, a process termed 'schwa absorption' by Shockey (2003, p. 22). For example, Shockey shows the realization of the second and third syllables of "people and" by a Southern Standard British speaker as having a syllabic [l] followed by a syllabic [n], so this is an instance of two consecutive syllables with no vowel. Low (2006) suggests that a duration of zero might be used in cases where there is no vowel that can be measured, but further reflection shows that this will not work. Firstly, if two consecutive vowels are assigned zero duration, then the average duration would be zero and so the PVI calculation would involve dividing by zero, which of course would result in a computational error; and secondly, if one syllable has no vowel so we assign it zero duration and then the next syllable has an extremely short vowel, say 10 ms, then the PVI value for these two syllables would be calculated as the difference, 10, divided by the average, 5, multiplied by 100, i.e., 200, which is an extremely high value. But both syllables are very short, so if we are attempting to evaluate whether there is an alternation of long and short vowels in the speech, then two consecutive short syllables ought to return a low value not a high one. Note that the PVI results are extremely sensitive to tiny changes in the measurement of short syllables, which is rather unfortunate.

The solution to this issue adopted here is to use a minimum of 30 ms for all measurements. As a result, if the vowel is very short, we do not worry about exact measurements, as the minimum value will be used instead.

The second issue involves the final syllable of each utterance. Firstly, measurement of an utterance-final vowel can be problematic, as it is often difficult to identify the end of a vowel in an open syllable in utterance-final position, especially in cases where the speech just seems to fade away rather than coming to a clear stop. But quite apart from practical considerations such as this, there is a more basic reason for omitting the measurement of the final syllable. It is well known that there is a tendency for the final syllable in an utterance to be lengthened in Singapore English (Low 2000), and this is almost certainly true in other varieties of English as well. The current study investigates Brunei English and the most widely spoken indigenous language of Brunei is Malay, which is similarly reported to have final syllable lengthening (Zuraidah et al. 2008). This probably has an effect on the English spoken in Brunei, so we are likely to encounter substantial utterance-final lengthening, which has a considerable impact on the results.

The solution to this issue is to omit the final syllable of the utterance from the calculations. It is entirely possible that utterances can consist of a series of evenly timed syllables with syllable-based rhythm followed by one final syllable that undergoes lengthening, and the omission of this final syllable will allow us to reflect this possibility accurately, though we should note that it means that there are limitations to the characterization of the speech that is being investigated as truly 'syllable-based'

and there is a further question about whether this proposed modification is appropriate for other languages.

The revised PVI formula can therefore be represented as:

$$PVI = 100 \times \left[\sum_{k=1}^{m-2} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| \Big/ (m-2) \right]$$

where d is the duration of the vowel in the *kth* syllable or 0.030 s, whichever is larger, and m is the number of syllables in the utterance.

The final issue concerns approximants. In a syllable such as *cold*, the dark /l/ is actually pronounced as a vowel by many speakers, including some RP speakers of British English, in a process known as L-vocalization (Wells 1982, p. 259), and this phenomenon is also reported to be widespread among speakers of other varieties of English, such as Singapore English (Tan 2005). If the /l/ is vocalized in this way, there really is no approximant that can be considered as part of the coda. In such cases, should all the vocalic part of the syllable be treated as the vowel? We could make a judgement based on whether there is perceived to be L-vocalization or not, but this would introduce a huge element of subjectivity to the measurements. In such situations, it is suggested that all the duration from the initial [k] up to the final [d] be treated as the vowel.

In the current study, the extracts of speech that are investigated involve no instances with /l/ as part of a consonant cluster, either at the start or end of the syllable. But a similar issue occurs with the /r/ in *from*. It is simply not possible to identify the end of the [r] and the start of the vowel in the majority of the tokens of this function word in the current data. Therefore, all the duration from the initial [f] to the final [m] was included in the measurements of this word.

The data that are investigated here include a token of *full*, which has a dark /l/ at the end, though this /l/ is not part of a consonant cluster. Issues involving this dark /l/ will be discussed below.

This third issue concerning the treatment of approximants that are part of a consonant cluster should be considered a little further. In the current study, the samples of speech being compared are varieties of English, so whatever decisions are made can be applied consistently to both sets of data. However, in studies involving the comparison of different languages, the decision about whether to include or exclude approximants in words such as *cold* and *from* in the measurement of vowels will have a huge effect on the results if one language allows consonant clusters but the other does not. In such a situation, the decision about how to deal with approximants needs to be reconsidered.

2.4 Subjects

The data investigated here is the read speech of 14 female and 6 male undergraduates at the University of Brunei Darussalam (UBD). The female speakers will be referred to with the 'F' prefix, while the male speakers have the 'M' prefix. All of them have good English, as English has been the medium of instruction in education

Table 2.1 Brunei speakers

Speaker	Age	Ethnicity	L1	L2	L3
F1	23	Malay	Malay	English	
F2	22	Malay	English	Malay	
F3	22	Malay	Malay	English	
F4	22	Chinese	English	Mandarin	Hokkien
F5	22	Chinese	English	Mandarin	Malay
F6	35	Malay	Malay	English	
F7	21	Malay	Malay	English	
F8	22	Malay	English	Malay	Tagalog
F9	22	Malay	Malay	English	
F10	22	Chinese	English	Malay	Mandarin
F11	20	Chinese/Dusun/Malay	Malay	English	Mandarin
F12	20	Malay	Indonesian	Malay	English
F13	21	Malay	Malay	English	
F14	21	Chinese	English	Malay	Mandarin
M1	20	Kedayan	Kedayan	English	Malay
M2	28	Iban	Iban	Malay	English
M3	23	Chinese	English	Hokkien	Mandarin
M4	21	Malay	Malay	English	
M5	22	Malay	English	Malay	Tagalog
M6	20	Malay	Malay	English	

in Brunei from the fourth year of primary school on since 1985 (Jones 2007), though some of the speakers state that their best language is Malay or a variety of Chinese. Details of the speakers and their languages are shown in Table 2.1.

If we assume that the rhythm of a variety of speech can be located along a syllable-/stress-based cline rather than belonging in one fixed category or another, then measurements using the PVI only really mean anything when they are compared with something else. For the purpose of comparison, therefore, the measurements of Brunei English are compared with similar measurements of the data of the three British male speakers whose vowels were plotted in Deterding (2006). They were aged 47, 48 and 57 at the time of the recordings and all three were lecturers at the National Institute of Education in Singapore.

2.5 Data

All the subjects were recorded using a high-quality microphone directly onto a computer using Praat software (Boersma and Weenink 2010). They read the Wolf passage, a text especially designed to facilitate the description of English because it has all the vowels and consonants of English in a range of environments (Deterding 2006). The full text of the Wolf passage is:

> There was once a poor shepherd boy who used to watch his flocks in the fields next to a dark forest near the foot of a mountain. One hot afternoon, he thought up a good plan to get some company for himself and also have a little fun. Raising his fist in the air, he ran

down to the village shouting "Wolf, Wolf." As soon as they heard him, the villagers all rushed from their homes, full of concern for his safety, and two of his cousins even stayed with him for a short while. This gave the boy so much pleasure that a few days later he tried exactly the same trick again, and once more he was successful. However, not long after, a wolf that had just escaped from the zoo was looking for a change from its usual diet of chicken and duck. So, overcoming its fear of being shot, it actually did come out from the forest and began to threaten the sheep. Racing down to the village, the boy of course cried out even louder than before. Unfortunately, as all the villagers were convinced that he was trying to fool them a third time, they told him, "Go away and don't bother us again." And so the wolf had a feast.

The original purpose of the Wolf passage was to provide an alternative to the North Wind and the Sun passage that has been used for many years as a standard text by the International Phonetic Association (IPA 1999, p. 39), and part of the rationale was that the North Wind and the Sun passage is not ideal for the measurement of rhythm because there are lots of instances of the approximant /w/ and also many sequences of full vowels rather than an alternation of strong and weak syllables. In fact, although the Wolf passage has been shown to be well suited for the description and measurement of the vowels and consonants of English, there remain many problems for measuring rhythm, with the approximant /w/ occurring regularly (*wolf* (×4), *was* (×4), *watch, one, with, while, once, were, away*), /j/ also occurring (*used, usual*) and also sequences of up to four strong syllables ("boy so much plea-" , "course cried out ev-") instead of the expected alternation of strong and weak syllables. Measurement here therefore just focuses on three phrases:

as soon as they heard him
full of concern for his safety
that had just escaped from the zoo

These three have lots of function words with potentially reduced vowels (*as, as, of, for, that, had, from, the*) as well as two bisyllabic content words with an unstressed first syllable (*concern*, *escaped*), so they allow us to focus on the effects of vowel reduction on the rhythm of speech.

However, as we will see, even this careful selection of just three utterances from the text leaves a number of problems with the measurement. While this means that there are some doubts about the validity of the results, the problems give us the opportunity to consider what we are actually measuring, how we should go about it, and what the output of the PVI in fact shows us.

2.6 Measurements

The vowel quality in the following syllables was evaluated auditorily: *as, as, of, con-*, *for, that, had, esc-*, *from, the*. Full vowels were annotated with '1' while reduced vowels were shown with '0'. For the first syllable of *escaped*, [e] was regarded as a full vowel while either [ɪ] or [ə] were considered reduced vowels.

For the rhythm of these three utterances, the duration of all the vowels except the last one was measured using Praat software (Boersma and Weenink 2010) and the PVI was calculated by means of Excel.

A major issue with regard to use of the PVI is reliability: how consistent are the measurements? One way to evaluate this is to repeat all the measurements after a period of time. In the current study, all measurements of the Brunei data were repeated after 6 months, so this offers an estimate of the degree of consistency. This only deals with intra-rater reliability and does not address the issue of inter-rater reliability, which is a major source of subjectivity in the measurement of rhythm using the PVI (Ong et al. 2005). In the current study, the comparative measurements of the Brunei and British data were all done by the same researcher, which helps ensure the consistency for judgements about the start and end of the vowels for all the data.

2.7 Results

The results of the perception of vowel quality are shown in Table 2.2.

These results show that the overwhelming majority of the function words *as, of, for, that, had* and *from* have a full vowel for the data of these Brunei speakers,

Table 2.2 Results of perception of the potentially reduced vowels for the Brunei speakers

Speaker	as	as	of	con-	for	that	had	esc-	from	the	Total
F1	1	1	0	0	1	1	1	1	1	0	7
F2	1	1	1	0	1	1	1	1	1	0	8
F3	1	0	1	0	0	1	1	0	0	0	4
F4	1	0	1	0	1	1	1	0	0	0	5
F5	1	1	1	1	1	1	1	0	1	0	8
F6	1	1	1	0	1	0	1	0	1	0	6
F7	1	1	1	0	1	1	1	0	1	0	7
F8	1	0	1	0	1	1	1	0	1	0	6
F9	1	1	1	0	1	1	1	1	1	0	8
F10	1	1	0	0	1	1	1	0	1	0	6
F11	1	1	1	1	1	1	1	0	1	0	8
F12	1	1	1	1	1	1	1	1	1	0	9
F13	1	1	1	0	0	1	1	1	0	0	6
F14	1	1	1	0	0	0	1	1	0	0	5
M1	1	1	0	0	1	1	1	0	1	0	6
M2	1	1	1	0	1	1	1	0	0	0	6
M3	1	1	1	0	1	1	1	1	1	0	8
M4	1	1	1	1	1	1	1	1	1	0	9
M5	1	1	0	0	1	1	1	0	0	0	5
M6	1	1	1	1	1	1	1	0	1	0	8
Total	20	17	16	5	17	18	20	8	14	0	

1 full vowel, *0* reduced vowel

Table 2.3 Results of the two attempts to measure the PVI of the Brunei speakers

Speaker	PVI (first)	PVI (second)	% Difference	Average	Full vowels
F1	35.60	39.12	9.9	37.36	7
F2	51.30	52.13	1.6	51.72	8
F3	42.10	47.50	12.8	44.80	4
F4	36.28	42.41	16.9	39.35	5
F5	37.06	38.29	3.3	37.68	8
F6	47.52	49.95	5.1	48.74	6
F7	45.33	46.33	2.2	45.83	7
F8	57.40	50.40	12.2	53.90	6
F9	18.78	21.59	15.0	20.19	8
F10	44.46	48.26	8.5	46.36	6
F11	49.37	48.64	1.5	49.01	8
F12	29.96	28.71	4.2	29.34	9
F13	43.53	49.21	13.0	46.37	6
F14	43.50	42.03	3.4	42.77	5
M1	41.56	42.09	1.3	41.83	6
M2	35.34	34.44	2.5	34.89	6
M3	37.00	32.57	12.0	34.79	8
M4	34.89	26.93	22.8	30.91	9
M5	40.13	39.91	0.5	40.02	5
M6	35.98	32.84	8.7	34.41	8
		Average	7.9	40.51	

The total of full vowels from Table 2.2 is shown in the last column

though it is important to note that all the speakers have reduced vowels in at least some words, as every token of *the* is produced with [ə]. The results also show that most of these speakers have [ə] in the first syllable of *concerned* and either [ɪ] or [ə] in the first syllable of *escaped*, so it seems that vowel reduction does generally occur in the unstressed syllables of polysyllabic words in these Brunei data, and the use of full vowels instead of reduced vowels is mostly found only in monosyllabic function words.

The results of the two attempts to measure the PVI are shown in Table 2.3, with the percentage difference between the two values shown in the % Difference column. In the final column, the total number of full vowels in the ten syllables that were analysed is repeated from Table 2.2, to facilitate an evaluation of the link between the use of full vowels and syllable-based rhythm.

There is reasonably good agreement between the two measurements of PVI, with an average of 7.9% difference between them. F9, F12 and M4 are the three speakers with the lowest PVI in both cases (suggesting the most syllable-based rhythm), while F2 and F8 have the highest values. However, there is also some degree of disagreement, the greatest discrepancies being a 22.8% difference for M4 and 16.9% disagreement for F4.

The PVI for the same three utterances from the Wolf passage read by the three British speakers was measured at 55.12, 61.79 and 58.38, with an average of 58.43, and there is a significant difference between this value and the overall average of

40.51 for the Brunei data (t = 3.62, df = 21, independent samples, two tailed, p = 0.0016). This suggests that Brunei English tends to have a more syllable-based rhythm than British English. However, we should note that there is substantial variation among the Brunei speakers, with F9 having a PVI value of 20.19 while F8 has a value of 53.90. Salbrina and Deterding (2010) also report considerable variability in the Brunei English of undergraduates at UBD, specifically with regard to the occurrence of rhoticity, and they suggest that Brunei English is not yet as well established as other regional varieties of English, in particular Singapore English.

Before we progress to consider these measurements in more detail and discuss what they really show, we can consider the relationship between vowel reduction and the results for PVI. Given that 1 shows a full vowel and 0 shows a reduced vowel, we would expect an inverse relationship between the PVI and the figure in the final column of Table 2.3, as lots of full vowels is expected to result in syllable-based rhythm. Indeed, the three speakers with the lowest PVI, F9, F12 and M4, all have lots of full vowels in their function words. However, there is a limit to this, as F2 and F8, the speakers who have the least syllable-based rhythm according to the PVI results, both have full vowels in nearly all their function words. The overall correlation between PVI and vowel reduction is −0.44, so this suggests that there exists a clear relationship between the two even though vowel reduction can only explain part of the variation in the PVI values.

2.8 Issues in the Measurement of the PVI

Despite the careful selection of a few suitable sentences from a read text, a number of difficulties persist in the measurements and these raise questions about the interpretation of the results. Do they really show the rhythm of this speech?

Firstly, there is a dark /l/ at the end of *full* and many speakers, including the Brunei speakers, vocalize this consonant. As a result, it can be extremely difficult to determine with any degree of certainty the duration of the vowel in *full* or the end of *full* and the start of *of*. For example, Fig. 2.1 is a spectrogram showing the pronunciation of "full of concern for his safety" by F10. In this extract, there is no indication of a consonant at the end of *full* and so it is difficult to decide in a consistent, principled way about the end of the [ʊ] vowel and the start of the consonant. In fact, about half of the tokens exhibit this problem.

In fact, in the same extract, there is often another issue. Although the [h] in *his* is reasonably clear in the utterance for F10 shown in Fig. 2.1, this is not always the case. Figure 2.2 is a spectrogram showing the pronunciation of the same utterance by M1. This speaker omits the [h] from *his*, as is quite standard in an unstressed function word in English (Roach 2009, p. 91) and, as a result, the two words *for* and *his* get merged together, so it is not clear where the boundary between the two vowels should be drawn. Furthermore, in measuring the duration of the vowels in this utterance, a subjective decision needs to be made in each case about whether there is an [h] or not in the data of each speaker, which is rather unfortunate.

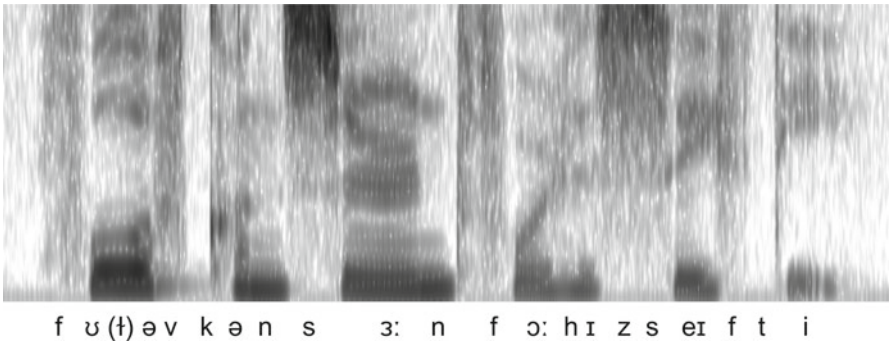


Fig. 2.1 Spectrogram of F10 saying "full of concern for his safety"

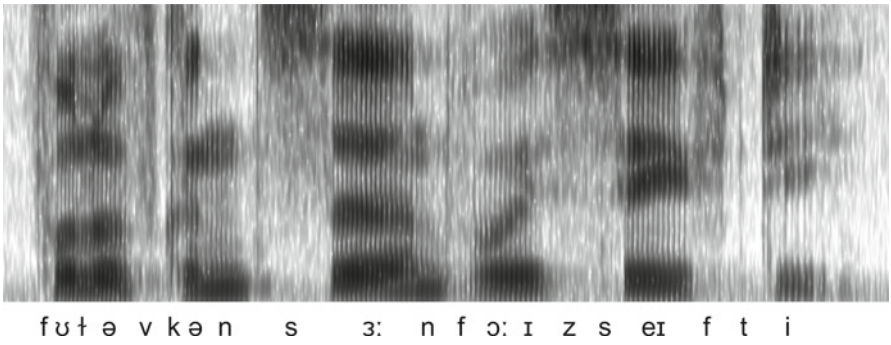


Fig. 2.2 Spectrogram of M1 saying "full of concern for his safety"

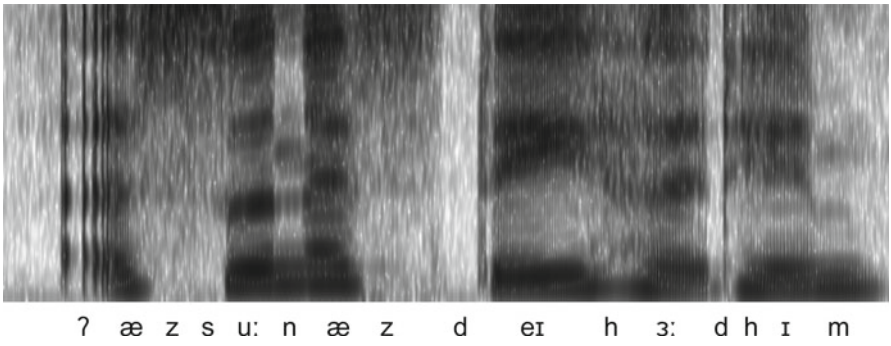


Fig. 2.3 Spectrogram of F5 saying "as soon as they heard him"

Such decisions about where to mark the boundary between two vowels can be a major source of subjectivity in these measurements; but other issues also occur. For example, a glottal stop sometimes occurs at the start of words beginning with a vowel, as illustrated by the spectrogram in Fig. 2.3, showing F5 saying "as soon as they heard him". There is a clear glottal stop before the start of the first token of *as*.

Should this be considered part of the vowel? Or maybe it is a kind of consonantal onset, so it is not part of the vowel.

These issues are methodological, affecting the implementation of the PVI measurements. But we should also consider what we are in fact measuring. All the measurements involve vowel duration, so are the results simply an indication of vowel reduction? Is rhythm really entirely dependent on vowel durations?

In reality, there are clear limitations to the extent to which the PVI results actually reflect the rhythm of speech. For example, F6 says "as soon as they heard him" with what seems like very deliberate, syllable-based rhythm but the PVI value for this utterance is 51.8, which is quite high, suggesting stress-based rhythm. The reason for this is that the vowel in *they* is rather long (149 ms), and it contrasts with the vowel in the preceding word *as* (73 ms) and also the following word *heard* (91 ms), resulting in a high PVI value. But perceptually, the utterance contains a whole sequence of full vowels, so it seems to have syllable-based rhythm. It seems that the PVI measurements are highly sensitive to measurements of the duration of long vowels as well as short ones, even though a sequence of long vowels is perceived as syllable-based rhythm.

These are issues to which we currently have no easy answers but which should be considered if the PVI is to continue to be used for the measurement of rhythm.

Finally, we might consider the teaching of rhythm. Low (2006) has suggested that this may be one practical application for the PVI, to provide a way of facilitating the speech training of learners of English who want to adopt stress-based rhythm for their speech. But should teachers be insisting on stress-based rhythm among their students? This issue will be considered in the next section.

2.9 Teaching Rhythm

Cruttenden (2008) insists that stress-based rhythm is essential for learners of English and many teaching textbooks treat the adoption of stress-timing as essential for fluency. For example, Teschner and Whitley (2004) introduce the metric foot in Chap. 1, considerably before the vowels and consonants of English are described in Chaps. 5 and 6, in the belief that stress-based rhythm and the alternation of strong and weak syllables is the fundamental framework on which the rest of the sound system of English is based. Similarly, Celce-Murcia et al. (1996, p. 26) note that the adoption of stress-based rhythm is "the most widely experienced pronunciation challenge for speakers of other languages", and they introduce a wide array of imaginative exercises, including chants, poems and jokes, to practice this kind of rhythm and thereby improve fluency (pp. 298–308).

But is stress-based rhythm really essential for English? Crystal (2003, p. 172) warns against imposing norms of rhythm where they are not appropriate and Jenkins (2007) excludes rhythm from the Lingua Franca Core (LFC), the set of pronunciation features which she suggests are essential for international intelligibility.

One issue that arises with regard to the relative absence of reduced vowels in styles of English that have a syllable-based rhythm is the effect it might have on psychological processes involved in perception. It has been suggested that speakers with reduced vowels in their function words tend to process these words differently from content words, so that for Inner-Circle Englishes the function words constitute the 'mortar' that holds together the 'bricks' of the content words (Field 2008). If some speakers of English use syllable-based rhythm, it is possible that they may process the language in a fundamentally different way from those with more stress-based rhythm.

The issue of teaching rhythm is likely to continue to be controversial and many teachers will remain convinced that the use of stress-based rhythm is vitally important for improving fluency among learners of English. But we might note that some exceptionally articulate speakers of English, such as Nelson Mandela and Kofi Annan, tend to have full vowels where other speakers would use reduced vowels and, as a result, the rhythm of their speech might be classified as substantially more syllable-based than that of most speakers from Inner-Circle countries. But nobody seems to suggest that there is anything wrong with their speech or that they should try to improve their intelligibility. So should we really be insisting on stress-based rhythm for the speech of learners? It is entirely possible that syllable-based rhythm actually enhances the intelligibility of English in many parts of the world (though perhaps not for most listeners in the UK or USA) and Crystal (1995) notes that the language of air traffic control ('Airspeak') tends to use "an even rhythm throughout" (p. 175) in order to achieve extra clarity. In conclusion, when travelling around the world or attending meetings with international participants, speakers might actually be encouraged to use syllable-based rhythm to ensure that they can be understood clearly and easily. Or at least, if they already have syllable-based rhythm in their English, there seems little reason to try to persuade them to abandon it.

2.10 Pragmatic Implications

It has here been suggested that it may not be necessary to teach stress-based rhythm, especially as syllable-based rhythm is probably more intelligible in many parts of the world. However, there is one final issue that should be considered. As mentioned above, Crystal (1995) reports that British English sometimes exhibits syllable-based rhythm to express emotions such as irritation and sarcasm. This raises two issues: if non-native speakers use syllable-based rhythm in their ordinary discourse, is it possible that their tone of voice might mistakenly be heard as carrying a hint of irritation or sarcasm by listeners in Britain? And secondly, if speakers from Outer-Circle countries are not aware of these implications, is it possible that they may misunderstand some of the essential pragmatic implications of the speech patterns adopted by people in Britain? Will they sometimes miss the fact that speakers are being sarcastic?

These are issues that need to be investigated further and it is probably true that people who intend to live in Britain for a long period of time need to be aware of

subtle shifts in the pragmatic undercurrents of the speech of the locals. However, in a world context, it seems less likely that people from Britain will misunderstand the pragmatic intentions of those who naturally use syllable-based rhythm and furthermore, speakers of British English would be foolhardy to expect listeners from around the world to detect such subtle emotional shifts in their tone of speech. Perhaps we can see here an important distinction between the ways English is spoken in a native context such as Britain or the USA and the ways it is used as an international language in international contexts.

References

Abercrombie, David. 1967. *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Barry, William J., and Michela Russo. 2003. Measuring rhythm. Is it separable from speech rate? In *Proceedings of the IP2003 Prosodic Interfaces Conference*, ed. Amina Mettouchi and Gaëlle Ferré, 15–20, Nantes, 27–29 Mar 2003.

Benton, Matthew, Liz Dockendorf, Wenhua Jin, Yang Liu, and Jerold A. Edmondson. 2007. The continuum of speech rhythm: Computational testing of speech rhythm of large corpora from natural Chinese and English Speech. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1269–1272, Saarbrücken, 6–10 Aug 2007.

Boersma, Paul, and David Weenink. 2010. Praat: Doing phonetics by computer. http://www.fon.hum.uva.nl/praat/. Accessed 19 Sept 2010.

Bond, Z.S., Dace Markus, and Verna Stockmal. 2003. Prosodic and rhythmic patterns produced by native and non-native speakers of a quantity-sensitive language. In *Proceedings of the 15th International Congress of Phonetic Sciences*, ed. M.J. Solé, D. Recasens and J. Romero, 526–530, Barcelona, 3–9 Aug 2003.

Brown, Adam. 1988. The staccato effect in the pronunciation of English in Malaysia and Singapore. In *New Englishes: The case of Singapore*, ed. Joseph Foley, 115–128. Singapore: Singapore University Press.

Celce-Murcia, Marianne, Donna M. Brinton, and Janet M. Goodwin. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.

Cruttenden, Alan. 2008. *Gimson's pronunciation of English*, 7th ed. London: Hodder Education.

Crystal, David. 1995. Documenting rhythmical change. In *Studies in general and English phonetics: Essays in honour of Professor J. D. O'Connor*, ed. Jack Windsor Lewis, 174–179. London: Routledge.

Crystal, David. 2003. *English as a global language*, 2nd ed. Cambridge: Cambridge University Press.

Cummins, Fred, and Robert Port. 1998. Rhythmic constraints on stress timing in English. *Journal of Phonetics* 26: 145–171.

Dankovičová, Jana, and Volker Dellwo. 2007. Czech speech rhythm and the rhythm class hypothesis. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1241–1244, Saarbrücken, 6–10 Aug 2007.

Deterding, David. 2001. The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics* 29: 217–230.

Deterding, David. 2006. The North Wind versus a Wolf: Short texts for the description and measurement of English pronunciation. *Journal of the International Phonetic Association* 36: 187–196.

Field, John. 2008. Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly* 42: 411–432.

Grabe, Esther, and Low Ee Ling. 2002. Durational variability in speech and the rhythm class hypothesis. In *Laboratory phonology 7*, ed. Carlos Gussenhoven and Natasha Warner, 515–546. Berlin: Mouton de Gruyter.

Hoequist, Charley E. 1983. The perceptual center and rhythm categories. *Language and Speech* 26: 367–376.

Hogg, Richard, and C.B. McCully. 1987. *Metrical phonology: A coursebook*. Cambridge: Cambridge University Press.

IPA. 1999. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press.

Jenkins, Jennifer. 2007. *English as a lingua franca: Attitude and identity*. Oxford: Oxford University Press.

Jones, Gary M. 2007. 20 years of bilingual education: Then and now. In *English in Southeast Asia: Varieties, literacies and literatures*, ed. David Prescott, 246–258. Newcastle: Cambridge Scholars Publishing.

Kachru, Braj B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle. In *English in the world: Teaching and learning the language and literatures*, ed. Randolph Quirk and Henry G. Widdowson, 11–30. Cambridge: Cambridge University Press.

Kachru, Braj B. 2006. *Asian Englishes: Beyond the canon*. Hong Kong: Hong Kong University Press.

Low Ee Ling. 2000. Is lexical stress placement different in Singapore English and British English? In *The English language in Singapore: Research on pronunciation*, ed. Adam Brown, David Deterding, and Low Ee Ling, 22–34. Singapore: Singapore Association for Applied Linguistics.

Low Ee Ling. 2006. A review of recent research on speech rhythm: Some insights for language acquisition, language disorders and language teaching. In *Spoken English, TESOL, and applied linguistics: Challenges for theory and practice*, ed. Rebecca Hughes, 99–125. New York: Palgrave Macmillan.

Low Ee Ling, Esther Grabe, and Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Journal of Phonetics* 29: 217–230.

Lleó, Conxita, Martin Rakow, and Margaret Kehoe. 2007. Acquiring rhythmically different languages in a bilingual context. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1545–1548, Saarbrücken, 6–10 Aug 2007.

Miller, M. 1984. On the perception of rhythm. *Journal of Phonetics* 12: 75–83.

Nolan, Francis, and Eva Liina Asu. 2009. The pairwise variability index and coexisting rhythms in language. *Phonetica* 66: 64–77.

Ong, Po Keng Fiona, David Deterding, and Low Ee Ling. 2005. Rhythm in Singapore English: A comparison of indexes. In *English in Singapore: Phonetic research on a corpus*, ed. David Deterding, Adam Brown, and Low Ee Ling, 74–85. Singapore: McGraw-Hill.

Pennycook, Alistair. 2007. *Global Englishes and transcultural flows*. London/New York: Routledge.

Ramus, Franck, Marina Nespor, and Jacques Mehler. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73(3): 265–292.

Roach, Peter. 1982. On the distinction between 'stress-timed' and 'syllable-timed' languages. In *Linguistic controversies*, ed. David Crystal, 73–79. London: Edward Arnold.

Roach, Peter. 2009. *English phonetics and phonology: A practical course*, 4th ed. Cambridge: Cambridge University Press.

Salbrina, Sharbawi, and David Deterding. 2010. Rhoticity in Brunei English. *English World-Wide* 31(2): 121–137.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Shockey, Linda. 2003. *Sound patterns of spoken English*. Malden: Blackwell.

Tan, Kah Keong. 2005. Vocalisation of /l/ in Singapore English. In *English in Singapore: Phonetic research based on a corpus*, ed. David Deterding, Adam Brown, and Low Ee Ling, 43–53. Singapore: McGraw-Hill.

Teschner, Richard V., and M.Stanley Whitley. 2004. *Pronouncing English: A stress-based approach with CD-ROM*. Washington, DC: Georgetown University Press.
Wells, J.C. 1982. *Accents of English*. Cambridge: Cambridge University Press.
Zuraidah, Mohd Don, Gerry Knowles, and Janet Yong. 2008. How words can be misleading: A study of syllable timing and "stress" in Malay. *The Linguistics Journal* 3(2). http://www.linguistics-journal.com/August_2008_zmd.php. Accessed 16 Sept 2010.

Chapter 3
Prosody and Second Language Teaching: Lessons from L2 Speech Perception and Production Research

Angelos Lengeris

3.1 Introduction

Pronunciation accuracy in a second language (L2) requires mastering production of both segmental (i.e., consonants and vowels) and suprasegmental or prosodic features of speech (i.e., features that extend over more than one segment such as lexical stress, pitch accent, rhythm and intonation) but teaching pronunciation of the latter is traditionally neglected in language classrooms. After the advent of the communicative approach to language teaching (e.g., Celce-Murcia et al. 1996; Morley 1991, 1994) that prioritized language function over language form, the study of L2 prosody has admittedly experienced an increasing interest among language teachers. In addition, following Pierrehumbert's (1980) pioneer work, research in intonation is one of the most fast-growing areas in linguistics with the autosegmental-metrical theory being the dominant framework in intonational research. Studies comparing the relative contribution of segmental *vs*. prosodic features of speech in degree of foreign accent have shown that deviations in the latter may affect listeners' judgement more than deviations in the former (e.g., Munro 1995; Munro and Derwing 1999; Derwing et al. 1998). Specifically, prosody has been found to be linked to accentedness, comprehensibility and intelligibility of speech (Anderson-Hsieh et al. 1992; Anderson-Hsieh and Venkatagiri 1994; Hahn 2004; Jilka 2000; Kang 2010; Kang et al. 2010; Munro and Derwing 2001; Pickering 2001; Trofimovich and Baker 2006). These findings are not surprising considering that prosody and intonation in particular plays a crucial role in communication by conveying not only linguistic information such as chunking the stream of speech in phrases, signalling new and contrastive information and disambiguating sentences that otherwise could sound ambiguous to the listener, but also paralinguistic information, i.e., information related to the

A. Lengeris (✉)
Department of Theoretical and Applied Linguistics, School of English, Aristotle University of Thessaloniki, Thessaloniki, Greece
e-mail: lengeris@enl.auth.gr

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_3,

identity, age, gender, and emotional state of the speaker. Misunderstandings due to the use of wrong intonation may even lead to negative evaluation and discrimination (e.g., Munro 2003).

Despite all this work showcasing the importance of prosody in L2 learning, its teaching is commonly ignored in the curriculum. A still popular view among teachers and learners holds that pronunciation and above all intonation cannot be taught, especially after the learner has passed what is considered to be the critical period for language acquisition. In addition, the majority of language teachers are non-native speakers of the target language and may lack the confidence or the ability to reproduce the prosodic patterns in a native-like manner. At the same time, L2 speech perception and production findings are usually disseminated only to academic audiences and do not reach the classroom. Even if they did, researchers and language practitioners do not necessarily share the same interests nor are research findings always presented in such a way as to facilitate implementation in the teaching curriculum. This article attempts to address these issues by reviewing important findings from L2 speech perception and production research indicating that (a) L2 learning difficulties are caused by native language (L1) experience and not because of a maturational-based loss in neural plasticity, which leaves the window for learning open well into adulthood and (b) the human brain can be re-trained to perceive and produce L2 segmentals and suprasegmentals using appropriate computer-based techniques developed and tested for their effectiveness in the laboratory over the last two decades.

3.2 Age and Second Language Learning

Young learners are better in acquiring an L2 than older learners. It is a common belief among teachers, policy makers and researchers since Lenneberg (1967) published the *Biological Foundations of Language* introducing the concept of a critical period for language acquisition that this decline in L2 performance is due to an age-related change in neural plasticity. It has therefore been claimed that biologically determined maturational constraints exist when learning the L2 grammar (Johnson and Newport 1989), syntax (Patkowski 1980) and pronunciation (Patkowski 1990). Age effects on L2 learning are reported in numerous studies examining the perception and production of vowels (e.g., Flege et al. 1999a) and consonants (e.g., Mackay et al. 2001). Studies concerned with the effect of age on the learning of L2 suprasegmentals are more limited compared to the segmental ones. The majority of these studies focus on degree of global foreign accent, a measure that combines segmental and suprasegmental aspects of speech (e.g., Flege et al. 1999b; Oyama 1976) confirming a decline in learners' performance with age; old learners are found to have stronger foreign accents than early learners. In a recent study, Huang and Jun (2009) examined the age effect on the acquisition of various aspects of American English prosody by Chinese learners. Three groups of Chinese learners participated, varying in their age of arrival in the United States while the length of residence in

the United States did not differ among groups. The study investigated Chinese learners' rate of English speech, the degree of foreign accent when speaking English (using low-pass filtered speech to remove segmental information while preserving the prosodic information) and the intonation patterns and prosodic groupings of their English speech production. The results confirmed an age effect on the acquisition of English prosody although the magnitude of the effect varied among the aspects of prosody that were examined.

The above studies demonstrate indisputable age effects on the acquisition of L2 segmentals and suprasegmentals. However, to support the view that such effects are due to an age-related loss in neural plasticity as proposed by the critical-period hypothesis, evidence is needed that (a) there is a sharp drop-off in the ability to learn a second language; (b) all early L2 learners achieve native-like performance; and (c) all late L2 learners fail to achieve native-like performance. On the contrary, a number of studies have shown that the perceptual system remains plastic enough to support learning well into adulthood and that there is no sharp drop-off in L2 learning ability but rather a gradual decline with age (Flege et al. 1999a, b). For example, Flege et al. (1999b) found that native Korean immigrants' degree of foreign accent increased as their age of arrival in the United States increased but there was no evidence of nonlinearity in Korean immigrants' performance. Further, other studies report that not all early bilinguals perform equally well (Flege et al. 1995, 1997) and that late bilinguals may achieve native-like pronunciation (e.g., Bongaerts et al. 1995, 1997; Moyer 1999). For example, Bongaerts et al. (1995, 1997) tested highly motivated Dutch learners of British English. None of the participants had received formal instruction in English before the age of around 12 and they were all exposed to a large amount of authentic L2 input delivered by native English speakers after entering the university. In the first study, foreign accent ratings were obtained for spontaneous speech, a text, 10 sentences and 25 words while in the second study ratings were obtained for 6 sentences. Bongaerts et al. (1995) found that all 10 Dutch participants were indistinguishable from native English speakers. Similarly, Bongaerts et al. (1997) found that 5 out of 11 participants met a criterion of 'native-likeness', i.e., their English sentence production received a mean rating that fell within 2 standard deviations of the mean rating obtained by English native speakers that were used as controls.

3.3 Linguistic Experience and Second Language Learning

If age-related changes in neural plasticity are not responsible for difficulties in learning the L2 segmentals and suprasegmentals, then what is it that makes it such a challenging task? Researchers believe that the advantage of early over late L2 learners is caused by our experience with our native language; as we grow up and acquire the sound system of our native language, our ability to learn patterns that are different from the native ones inevitably declines. A change in infants' perceptual abilities during the first year of life has been extensively described in a number of

studies conducted the past 30 years. It has therefore been shown that, in the early months of life, infants are able to discriminate all sounds that are used to signal contrasts in any language (Aslin et al. 1981; Eimas et al. 1971; Trehub 1976). However, by the end of their first year infants fail to discriminate non-native consonant contrasts (Werker et al. 1981; Werker and Tees 1983, 1984). Sensitivity to non-native vowel contrasts appears to decline somewhat earlier, at around 6 months of age (Kuhl et al. 1992; Polka and Werker 1994). During their first year of life, infants demonstrate a similar perceptual reorganization for suprasegmental features of speech such as rhythm (Jusczyk et al. 1993) and lexical tone (Mattock and Burnham 2006, Mattock et al. 2008). For example, Jusczyk et al. (1993) showed that 9-month-old American infants, in contrast to 6-month-old American infants prefer to listen to words with a strong/weak stress pattern, which is the most frequently used pattern in English over words with a weak/strong pattern, indicating that experience with the prosodic features of the ambient language affects infants' response to language.

The role of L1 'tuning' in L2 speech learning is discussed in current L2/cross-linguistic models, such as the Perceptual Assimilation Model (Best 1995; Best and Tyler 2007), the Speech Learning Model (Flege 1995, 2002), and the Native Language Magnet model (Kuhl et al. 1992, 2008; Kuhl 2000). All three models agree that L1 language experience interferes with L2 learning and that the relationship between the L1 and L2 sound inventories can predict whether or not a specific L2 sound (or a specific L2 contrast) will pose difficulty to the learner. For example, the Speech Learning Model posits that as the L1 categories develop with age they become more powerful attractors of the L2 categories (e.g., Walley and Flege 1999). Several studies have demonstrated the role of linguistic experience in learning the L2 vowels (e.g., Cebrian 2006; Flege et al. 1999a; Flege and MacKay 2004; Iverson and Evans 2007; Lengeris 2009; Polka 1995) and consonants (e.g., Best et al. 2001; Guion et al. 2000; Hattori and Iverson 2009; Iverson et al. 2003; Mackay et al. 2001). For example, Spanish and Greek learners of English show a very poor discrimination of the English tense-lax vowel contrast /iː/-/ɪ/ because they lack such a contrast in their L1, having a single vowel category in the F1/F2 vowel space occupied by the two English vowels (Cebrian 2006 for Spanish learners; Lengeris 2009 for Greek learners). Likewise, Japanese speakers are very poor at differentiating English /ɹ/ from /l/ because they pay attention to the non-critical second formant frequency (which is important for the perception of the Japanese voiced tap /ɾ/) instead of the critical third formant frequency (Iverson et al. 2003).

Effects of linguistic experience are also reported in studies on suprasegmental features of speech, specifically on the acquisition of stress (e.g., Archibald 1993; Dupoux et al. 1997, 2001; Guion et al. 2004; Peperkamp and Dupoux 2002; Peperkamp et al. 2010; Yu and Andruski 2010) and tone (e.g., Gottfried and Suiter 1997; Hallé et al. 2004; So and Best 2010; Wayland and Guion 2004). Clear effects of L1 experience on the way learners perceive and produce the L2 intonational patterns are also reported. Early studies focused on the errors produced by learners (e.g., Backman 1979; Willems 1982) but contemporary research has acknowledged the need to adopt a generally agreed framework for intonational analysis to better

examine cross-linguistic similarities and differences in intonation. Mennen (2006) discusses the difficulties in comparing the findings of different studies in intonation research, as well as the potential of the Autosegmental framework of intonational analysis (Pierrehumbert 1980) in investigating L2 intonation. The model distinguishes between the underlying phonological representation of intonation (e.g., tonal inventory) and its phonetic manifestation (e.g., F0 peak alignment), providing a test-bed for the acquisition of L2 intonational targets and their phonetic realization. It is therefore not surprising that a growing number of studies have begun using the Autosegmental framework to examine the influence of L1 on the learning of L2 intonation during the last decade (e.g., Atterer and Ladd 2004; Jilka 2000; Mennen 2004, 2006).

Mennen (2004) investigated the extent to which the L1 intonation system can exert an influence on the acquisition of L2 intonation at the phonetic level. The study examined the production of Greek pre-nuclear rises by advanced Dutch learners of Greek. Greek and Dutch use phonologically identical pre-nuclear rises in declarative sentences but there are cross-linguistic differences in the phonetic manifestation of the rise. In Greek, the alignment of the peak is realized in the vowel following the accented syllable, whereas in Dutch the peak is realized slightly earlier, within the accented syllable. Furthermore, in Dutch the alignment of the peak is affected by the length of the vowel of the accented syllable (i.e., earlier when the vowel is long and later when the vowel is short), whereas in Greek it is not (there are no short-long distinctions in Greek). Five Dutch learners of Greek, all teachers of Greek at University level, participated in the study. The production of pre-nuclear rises by a group of Dutch native speakers and a group of Greek native speakers with no knowledge of Greek and Dutch respectively, were recorded for control reasons. Mennen (2004) found that four out of five Dutch learners of Greek transferred their L1 (Dutch) phonetic realization of pre-nuclear rises when speaking Greek (i.e., they aligned the peak earlier than Greek speakers) and only one Dutch learner managed to show native-like performance. Interestingly, Mennen (2004) reports a bi-directional interference in the production of pre-nuclear rises by those four Dutch learners of Greek; not only did they differ from Greek controls in their production of L2 (Greek) intonation but they also differed from Dutch controls in their production of L1 (Dutch) intonation. Only one Dutch learner managed to achieve native-like performance in peak alignment in both languages.

3.4 Learning in Naturalistic and Formal Settings

The role of L2 experience – usually indexed by the length of residence (LOR) in an L2 setting – in the acquisition of the L2 segmentals and suprasegmentals has been extensively examined in the literature but there are inconsistencies in findings across studies (see Piske et al. 2001 for a review of factors affecting degree of foreign accent in an L2). Some studies have found evidence supporting the importance of experience on L2 learning (e.g., Asher and Garcia 1969; Flege and Fletcher 1992;

Flege et al. 1997) while others report no such effect (e.g., Cebrian 2006; Oyama 1976; Piper and Cansin 1988). Trofimovich and Baker (2006) studied the effect of experience on the production of five English suprasegmentals (stress timing, peak alignment, speech rate, pause frequency and pause duration) by three groups of Korean learners of English who had been immersed in the United States after puberty and differed in their length of immersion (3 months, 3 and 10 years). Participants performed a delayed repetition task (declaratives as responses to question prompts). The sentences produced by Korean speakers were low-pass filtered to remove segmental effects and were rated by native English speakers for degree of global foreign accent. The sentences were also measured acoustically in terms of the five target suprasegmentals. All Korean speakers were found to be more accented than a control group of English speakers but those who were less experienced (i.e., 3 months of residence) were more accented than those with 3 and 10 years of residency (but there was no difference between the last two groups). The acoustic analysis showed that the amount of L2 experience correlated to stress timing but not to the other four suprasegmentals tested. In a following study, Trofimovich and Baker (2007) studied the effect of experience on the production of the same five English suprasegmentals by Korean learners of English who had been immersed in the United States before puberty and differed in their L2 experience (1 *vs*. 11 years of residency). The results showed that the latter group outperformed the former in all 5 suprasegmentals as well as in degree of global foreign accent. Those Koreans with 11 years of residency achieved native-like levels of performance in the global foreign accent task and in four suprasegmentals (all except speech rate).

Flege and Liu (2001) suggested that the lack of an effect of LOR in some studies may have been due to the quality of the L2 input the sampling population received. In their study, Flege and Liu (2001) examined the effect of LOR on L2 learning by means of a consonant identification task, a grammaticality judgment task and a listening comprehension task. The participants were adult Chinese speakers who had lived in the United States from 0.5 to 3.8 years (short LOR group) and from 3.9 to 15 years (long LOR group). Half of the participants in each group were university students while the remaining participants had worked full-time during their stay in the US. In all three tasks, an effect of LOR was found for the group of students but not for the non-students; only the former group achieved higher scores following immersion, a finding which demonstrates that L2 learning depends on the quality of native-speaker input that the learner receives (the two groups did not differ in terms of self-reported percentage use of English). Flege (2009) further discussed the importance of input in L2 learning. According to the author, both quality and quantity of input are important; residence in a foreign country is likely to be beneficial only for immigrants who receive a sufficient amount of L2 input via interaction with native speakers, especially via participation in social activities. In cases where immigrants receive a greater amount of L1-accented input than authentic input, the amount of L2 experience cannot be a reliable predictor of success in L2 learning.

Indirect evidence for the importance of authentic input when learning a second language comes from research in formal language settings (Elliott 1995; Fullana and MacKay 2010; Gallardo del Puerto et al. 2005; García Lecumberri and

Gallardo del Puerto 2003; MacKay and Fullana 2007; Mora and Fullana 2007). These studies report no effect of language instruction on L2 perception and production, which indicates that L2 exposure at a young age and several years of formal instruction may not lead to better pronunciation (see Singleton and Ryan 2004 for a review). This can be explained by the fact that classroom instruction is normally limited to a few hours per week, focuses on L2 form and the teacher who acts as a model to students delivers, in most cases, L1-accented input.

3.5 Learning in Laboratory Settings

One of the strongest arguments against the view that there is an age-related loss in neural plasticity comes from a number of computer-based training studies conducted over the past years. These studies have consistently shown that adults from various language backgrounds can be retrained to hear and produce L2 segmentals and suprasegmentals using structured intensive training procedures. Early studies attempting to modify perception of sounds adopted discrimination training whereby the trainees hear two synthetic stimuli in each trial and are asked to decide whether the two stimuli are identical or different (e.g., Carney et al. 1977; Pisoni et al. 1982; Strange and Dittmann 1984). These early studies showed a learning effect on the training stimuli but no transfer of learning to natural tokens. This was attributed partly to the use of discrimination training and partly to the low variability of the training stimuli. Regarding the former issue, it has been claimed that discrimination training tends to tailor learners' attention to within-category differences instead of focusing on between-category differences that are crucial for categorization. Regarding the latter issue, it is believed that the use of a single talker and context impedes transfer of learning to other talkers and contexts.

An alternative approach to training that has dominated the field over the past 20 years is the high-variability phonetic training technique. As its name indicates, this method emphasizes the importance of exposure to natural minimal pairs contrasting the target sounds in multiple environments spoken by multiple talkers thus resembling real-world communication with native speakers of the target language. High-variability training consists of an identification task with feedback whereby the trainees hear a single stimulus in each trial and are asked to label the sound using a number of given L2 categories. This approach to training has been found to significantly improve by about 20% the perception of L2 consonants and vowels (e.g., Logan et al. 1991; Lively et al. 1994; Hazan et al. 2005; Iverson and Evans 2009; Lengeris and Hazan 2010; Nishi and Kewley-Port 2007, 2008). Importantly, perceptual improvement is not limited to the stimuli used in training but generalizes to new words containing the target sounds and to talkers that were not heard during training. Perceptual learning is retained several months after training (Lively et al. 1994; Bradlow et al. 1999) and transfers to speech-in-noise vowel perception (Lengeris and Hazan 2010) and to consonant (Bradlow et al. 1997; Hazan et al. 2005) and vowel production (Lambacher et al. 2005; Lengeris and Hazan 2010).

Among the L2 segmentals that have been trained are the English vowel distinctions for Spanish, Greek, German and French speakers, the English /r/-/l/ distinction for Japanese speakers, the English word final /t/-/d/ distinction for Chinese speakers and the Hindi dental-retroflex stop distinction for English and Japanese speakers. Note that consonant studies usually train binary L2 contrasts (e.g., /r/-/l/, or /t/-/d/) whereas vowel studies have successfully trained several L2 vowels at the same time.

Wang and Munro (2004) examined whether computer-based training procedures such as these reviewed so far can be effective in improving L2 vowel perception in a classroom setting whereby learners are given some control over training. The authors trained Mandarin and Cantonese speakers on three English vowel contrasts /iː/-/ɪ/, /e/-/æ/, and /ʊ/-/uː/ using synthetic and natural vowel stimuli. The trainees were asked to schedule their 50–60 min training sessions (2–3 per week) and in each session they could recycle the training blocks as desired. After training, participants improved their perception of English vowels; their improvement was transferred to a new context and was retained 3 months after training. The results of this study are important because they show that research knowledge from laboratory training studies can be put into practice in a classroom environment.

Another line of research has been exploring the implementation of fundamental frequency visualization software for teaching L2 intonation since the 1970s (Abberton and Fourcin 1975; Anderson-Hsieh 1992, 1994; Chun 1998; De Bot 1983; Hardison 2004; Levis and Pickering 2004; Spaai and Hermes 1993; Taniguchi and Abberton 1999; Weltens and de Bot 1984). Early software (and hardware) used to be expensive and difficult to use but nowadays computer technology for speech analysis is becoming widely accessible. A number of analysis programs developed by the research community such as Praat (Boersma and Weenink 2009), SFS (Huckvale 2008), Wavesurfer (Sjölander and Beskow 2000) and Speech Analyzer (SIL 2007) are freely available online. Learners can record themselves and see on the screen a visual representation of the pitch contour of their speech. The visualization of pitch is relatively easy to interpret by inexperienced learners and does not require extensive phonetic knowledge, as is required for the interpretation of e.g., spectrograms or waveforms. Learners can also compare their production both auditory and visually with a model speaker, which helps to raise learners' awareness of differences between L1 and L2 intonation patterns.

Anderson-Hsieh (1992) trained international teaching assistants (mainly Chinese and Korean native speakers) on English word stress, rhythm, linking and intonation using visual feedback. Although the paper's aim was to demonstrate how visual feedback can be used as a tool for teaching suprasegmentals rather than evaluating the effectiveness of training using statistical analyses, the author reports that training had a positive effect on the learners' performance and attributes this finding to the fact that a visual representation of suprasegmentals in real time accompanied the auditory feedback provided to the learners. De Bot (1983) compared the effects of audio-visual feedback and audio-only feedback on Dutch students' learning of English intonation. The results showed that the group of students who received audio-visual feedback improved more than those who received audio-only feedback, as judged by three teachers of English. Taniguchi and Abberton (1999) examined

the effectiveness of interactive visual feedback on Japanese speakers' production of English intonation. The Japanese speakers were 12 college students who attended the UCL Summer Course in English Phonetics. All students attended lectures and practical lessons in English intonation but only half of them received training with visual feedback during their intonation lessons. All 12 Japanese speakers improved their English intonation after 2 weeks but those who received interactive visual feedback outperformed those who did not receive visual feedback.

Hardison (2004) examined the effectiveness of computer-assisted training on the acquisition of French prosody by native speakers of American English using pitch displays and multiple recordings spoken by multiple native French speakers as feedback. This training protocol differs from protocols used in suprasegmental training studies reviewed so far, as native speakers of the target language were used as feedback and not as models to imitate. Hardison's study consisted of two experiments. In Experiment 1, 26 American English speakers, all undergraduate university students, participated. Of those speakers, 16 were trained while the remaining speakers served as controls, i.e., performed only the pre/post test without receiving any training. In the pre/post test, participants were asked to produce 20 French sentences while another 20 novel sentences in the post-test evaluated generalization of learning. Training consisted of thirteen 40-min sessions with sets of 30 sentences spoken by three native French speakers. During each training session, participants were asked to produce one set of 30 sentences. The pitch contours of their sentences were displayed in real time on the screen. A French speaker's version of that sentence was displayed on the screen in a different colour and played out aloud. Three native French speakers evaluated the recordings made by the American English speakers on a 7-point scale. The results indicated that the trainees improved their production and that this improvement generalized to the set of novel sentences. Experiment 2 used a memory recall task using filtered stimuli that preserved prosodic information while reducing segmental information to examine whether learners were able to identify the exact lexical content of the training sentences based on prosodic information alone. The trainees succeeded about 80% of the time in doing so, demonstrating according to the author that prosodic and lexical information is stored together in memory traces. Finally, participants' responses from questionnaires indicated a greater awareness of the various aspects of speech after training and increased confidence when speaking French.

Following the growing interest in discourse intonation among intonation experts (e.g., Bolinger 1989; Brazil 1997; Chun 2002) and the success of intonation training studies that have used sentence-level training materials, researchers have begun to explore ways to teach intonation patterns that occur in communicative contexts (Chun 1998; Levis and Pickering 2004). For example, Hardison (2005) examined whether computer-based training can improve the production of L2 prosody at the level of contextualized speech, using pitch displays and discourse-level training materials. Twenty-eight Chinese advanced learners of English participated in the study. Half of the participants were trained on discourse-level materials and half were trained on individual sentences. Three native English speakers provided global prosody ratings for both groups of Chinese speakers. The results showed that all

speakers showed transfer of learning to natural discourse as a result of training and that the group of speakers that received contextualized input improved more than the group that received sentence-level input. Prosody ratings of speech materials produced by the same Chinese learners 1 week after training revealed sustained improvement. These results showcase that computer technology can be effective in teaching not only the typical sentence-type intonation patterns (e.g., declaratives, wh-questions, yes-no questions, etc.) but also discourse-level intonation patterns.

3.6 Conclusion

The acquisition of second language prosody is undeniably a difficult task. However, a growing body of experimental work supports the view that, as with segmental aspects of speech, the window of learning is not closed even in adulthood. This is because difficulties in L2 learning are mainly caused by native language experience and not because of an age-related change in neural plasticity that would make learning unfeasible. Strong evidence for this plasticity comes from laboratory studies showing that the adult brain can be trained to hear non-native differences by using appropriate computer-based training techniques. What is even more encouraging is that perceptual training can improve not only the perception of L2 segmentals and suprasegmentals but also their production. The most successful training protocols adopt the so-called high-variability approach, which has been proven to create robust and long-lasting new categories. In line with exemplar and statistical models of speech perception (Goldinger 1996; Johnson 1997; Pierrehumbert 2002; Maye et al. 2002), in order to create new categories we simply need to expose the learner to multiple natural tokens of the target sounds produced by various speakers. Exposure to authentic and variable L1 input is therefore vital in the learning process. In the case of suprasegmentals, the visualization of the pitch contour can help those learners who cannot rely only on their own perception of intonation to improve their production of both sentence- and discourse-level prosody. The area of L2 intonation training thus provides a perfect example of how valuable insights from academic work can be put into practice to assist the teaching of one of the most difficult aspects of second language pedagogy.

References

Abberton, E., and A. Fourcin. 1975. Visual feedback and the acquisition of intonation. In *Foundations of language development*, ed. E.H. Lenneberg and E. Lenneberg, 157–165. New York: Academic Press.

Anderson-Hsieh, J. 1992. Using electronic visual feedback to teach suprasegmentals. *System* 20: 51–62.

Anderson-Hsieh, J. 1994. Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal* 11: 5–22.

Anderson-Hsieh, J., and H. Venkatagiri. 1994. Syllable duration and pausing in the speech of intermediate and high proficiency Chinese ESL speakers. *TESOL Quarterly* 28: 807–812.
Anderson-Hsieh, J., R. Johnson, and K. Koehler. 1992. The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning* 42: 529–555.
Archibald, J. 1993. The learnability of English metrical parameters by adult Spanish speakers. *International Review of Applied Linguistics* 31: 129–141.
Asher, J., and R. Garcia. 1969. The optimal age to learn a foreign language. *The Modern Language Journal* 53: 334–341.
Aslin, R.N., D.B. Pisoni, B.L. Hennessy, and A.J. Perey. 1981. Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development* 52: 1135–1145.
Atterer, M., and D.R. Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics* 32: 177–197.
Backman, N.E. 1979. Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin* 4(2): 239–266.
Best, C.T. 1995. A direct realist view of cross-language speech perception. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. W. Strange, 171–204. Baltimore: York Press.
Best, C.T., and M.D. Tyler. 2007. Non-native and second-language speech perception: Commonalities and complementarities. In *Second language speech learning*, ed. M. Munro and O.-S. Bohn, 13–34. Amsterdam: John Benjamins.
Best, C.T., G.W. McRoberts, and E. Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America* 109: 775–794.
Boersma, P., and D. Weenink. 2009. Praat: Doing phonetics by computer. [Computer Program]. http://www.praat.org/. Accessed 12 Aug 2009.
Bolinger, D. 1989. *Intonation and its uses: Melody in grammar and discourse*. London: Arnold.
Bongaerts, T., B. Planken, and E. Schils. 1995. Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In *The age factor in second language acquisition*, ed. D. Singleton and Z. Lengyel, 30–50. Clevedon: Multilingual Matters.
Bongaerts, T., C. Van Summeren, B. Planken, and E. Schils. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–465.
Bradlow, A.R., D.B. Pisoni, R. Akahane-Yamada, and Y. Tohkura. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101: 2299–2310.
Bradlow, A.R., R. Akahane-Yamada, D.B. Pisoni, and Y. Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics* 61: 977–985.
Brazil, D. 1997. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.
Carney, A.E., G.P. Widin, and N.F. Viemeister. 1977. Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America* 62: 961–970.
Cebrian, J. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34: 372–387.
Celce-Murcia, M., D. Brinton, and J. Goodwin. 1996. *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge: Cambridge University Press.
Chun, D.M. 1998. Signal analysis software for teaching discourse intonation. *Language Learning and Technology* 2: 61–77. http://llt.msu.edu/vol2num1/article4. Accessed 11 Mar 2009.
Chun, D.M. 2002. *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.
De Bot, K. 1983. Visual feedback on intonation I: Effectiveness and induced practice behaviour. *Language and Speech* 26: 331–350.

Derwing, T.M., M.J. Munro, and G. Wiebe. 1998. Evidence in favour of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.
Dupoux, E., C. Pallier, N. Sebastian, and J. Mehler. 1997. A destressing "deafness" in French? *Journal of Memory and Language* 36: 406–421.
Dupoux, E., S. Peperkamp, and N. Sebastian. 2001. A robust method to study stress "deafness". *Journal of the Acoustical Society of America* 110: 1606–1618.
Eimas, P.D., E.R. Siqueland, P. Jusczyk, and J. Vigorito. 1971. Speech perception in infants. *Science* 171: 303–306.
Elliott, A.R. 1995. Foreign language phonology: Field independence, attitude and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal* 79: 530–542.
Flege, J.E. 1995. Second language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Issues in cross-language research*, ed. W. Strange, 233–277. Baltimore: York Press.
Flege, J.E. 2002. Interactions between the native and second-language phonetic systems. In *An integrated view of language development: Papers in honor of Henning Wode*, ed. P. Burmeister, T. Piske, and A. Rohde, 217–244. Trier: Wissenschaftlicher Verlag.
Flege, J.E. 2009. Give input a chance! In *Input matters in SLA*, ed. T. Piske and M. Young-Scholten, 175–190. Bristol: Multilingual Matters.
Flege, J.E., and K.L. Fletcher. 1992. Talker and listener effects on the degree of perceived foreign accent. *Journal of the Acoustical Society of America* 91(1): 370–389.
Flege, J.E., and S. Liu. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–552.
Flege, J.E., and I.R.A. Mackay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26: 1–34.
Flege, J.E., M. Munro, and I.R.A. MacKay. 1995. Effects of age of second language learning on the production of English consonants. *Speech Communication* 16: 1–26.
Flege, J.E., E.M. Frieda, and T. Nozawa. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–186.
Flege, J.E., I.R.A. MacKay, and D. Meador. 1999a. Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America* 106: 2973–2987.
Flege, J.E., G.H. Yeni-Komshian, and S. Liu. 1999b. Age constraints on second language acquisition. *Journal of Memory and Language* 41: 78–104.
Fullana, N., and I.R.A. MacKay. 2010. L2 fluency in the production of English extemporaneous speech by Catalan/Spanish bilinguals. In *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech*, 125–130, Poznan, Poland. http://ifa.amu.edu.pl/newsounds/. Accessed 5 Jan 2011.
Gallardo del Puerto, F., M.L. García Lecumberri, and J. Cenoz. 2005. Degree of foreign accent and age of onset in formal school instruction. http://www.phon.ucl.ac.uk/home/johnm/ptlc2005/pdf/ptlcp48.pdf. Accessed 2 Feb 2009.
García Lecumberri, M.L., and F. Gallardo del Puerto. 2003. English FL sounds in school learners of different ages. In *Age and the acquisition of English as a foreign language*, ed. M.P. García Mayo and M.L. García Lecumberri, 115–135. Clevedon: Multilingual Matters.
Goldinger, S.D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22: 1166–1183.
Gottfried, T., and T. Suiter. 1997. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics* 25: 207–231.
Guion, S.G., J.E. Flege, R. Akahane-Yamada, and J.C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107: 2711–2724.
Guion, S.G., T. Harada, and J.J. Clark. 2004. Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition* 7: 207–226.
Hahn, L.D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38: 201–223.

Hallé, P., Y.-C. Chang, and C.T. Best. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32: 395–421.

Hardison, D.M. 2004. Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology* 8: 34–52. http://llt.msu.edu/vol8num1/pdf/hardison.pdf. Accessed 22 Oct 2008).

Hardison, D.M. 2005. Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal* 22(2): 175–190.

Hattori, K., and P. Iverson. 2009. English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America* 125: 469–479.

Hazan, V., A. Sennema, M. Iba, and A. Faulkner. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47: 360–378.

Huang, B., and S. Jun. 2009. Age effect on the acquisition of second language prosody. In *Proceedings of the 33rd Boston University Conference on Language Development.* http://www.bu.edu/linguistics/APPLIED/BUCLD/supplement33/Huang.pdf. Accessed 6 Sept 2010.

Huckvale, M. 2008. Speech Filing System tools for speech research, University College London. [Computer program]. http://www.phon.ucl.ac.uk/resource/sfs/. Accessed 3 Apr 2009.

Iverson, P., and B.G. Evans. 2007. Learning English vowels with different first language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America* 122: 2842–2854.

Iverson, P., and B.G. Evans. 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America* 126: 866–877.

Iverson, P., P.K. Kuhl, R. Akahane-Yamada, E. Diesch, Y. Tohkura, and A. Kettermann. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: 47–57.

Jilka, M. 2000. The contribution of intonation to the perception of foreign accent. PhD thesis, University of Stuttgart, Stuttgart.

Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In *Talker variability in speech processing*, ed. K. Johnson and J. Mullenix, 145–166. San Diego: Academic.

Johnson, J., and E. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21: 60–99.

Jusczyk, P.W., A. Cutler, and N.J. Redanz. 1993. Infants' preference for the predominant stress patterns of English words. *Child Development* 64: 675–687.

Kang, O. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38: 301–315.

Kang, O., D. Rubin, and L. Pickering. 2010. Suprasegmental measures of accentedness and judgments of English language learner proficiency in oral English. *Modern Language Journal* 94: 554–566.

Kuhl, P.K. 2000. A new view of language acquisition. *National Academy of Sciences USA* 97: 11850–11857.

Kuhl, P.K., K.A. Williams, F. Lacerda, K.N. Stevens, and B. Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255: 606–608.

Kuhl, P.K., B.T. Conboy, S. Coffey-Corina, D. Padden, M. Rivera-Gaxiola, and T. Nelson. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded NLM-e. *Philosophical Transactions of the Royal Society* B363: 979–1000.

Lambacher, S.G., W.L. Martens, K. Kakehi, C.A. Marasinghe, and G. Molholt. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied PsychoLinguistics* 26: 227–247.

Lengeris, A. 2009. Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica* 66: 169–187.

Lengeris, A., and V. Hazan. 2010. The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *Journal of the Acoustical Society of America* 128(6): 3757–3768.

Lenneberg, E.H. 1967. *Biological foundations of language*. New York: Wiley.

Levis, J., and L. Pickering. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32(4): 505–524.

Lively, S.E., D.B. Pisoni, R. Yamada, Y. Tohkura, and T. Yamada. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America* 96: 2076–2087.

Logan, J.S., S.E. Lively, and D.B. Pisoni. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America* 89: 874–886.

Mackay, I.R.A, and N. Fullana. 2007. Starting age and exposure effects on EFL learners' sound production in a formal learning context. In *Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*, 324–335, Florianópolis, Brazil. http://www.nupffale.ufsc.br/newsounds/Papers/28.Mackay_Fullana.pdf. Accessed 11 May 2009.

Mackay, I.R.A., D. Meador, and J.E. Flege. 2001. The identification of English consonants by native speakers of Italian. *Phonetica* 58: 103–125.

Mattock, K., and D. Burnham. 2006. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy* 103: 241–264.

Mattock, K., M. Molnar, L. Polka, and D. Burnham. 2008. The developmental course of lexical tone perception in the first year of life. *Cognition* 106(3): 1367–1381.

Maye, J., J.F. Werker, and L. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101–B111.

Mennen, I. 2004. Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–563.

Mennen, I. 2006. Phonetic and phonological influences in non-native intonation: An overview for language teachers. *QMUC Speech Science Research Centre Working Papers* WP-9.

Mora, J.C., and Fullana, N. 2007. Production and perception of /i□/-/□/ and /æ/-/□/ in a formal setting: Investigating the effects of experience and starting age. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1613–1616, Saarbrücken, Germany.

Morley, J. 1991. The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly* 25(3): 481–520.

Morley, J. 1994. A multidimensional curriculum design for speech/pronunciation instruction. In *Pronunciation theory and pedagogy: New views, new directions*, ed. J. Morley, 64–91. Alexandria: TESOL Publications.

Moyer, A. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21: 81–108.

Munro, M.J. 1995. Non-segmental factors in foreign accent. *Studies in Second Language Acquisition* 17: 17–34.

Munro, M.J. 2003. A primer on accent discrimination in the Canadian context. *TESL Canada Journal* 202: 38–51.

Munro, M.J., and T. Derwing. 1999. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 49(Supplement 1): 285–310.

Munro, M.J., and T.M. Derwing. 2001. Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition* 23: 451–468.

Nishi, K., and D. Kewley-Port. 2007. Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research* 50: 1496–1509.

Nishi, K., and D. Kewley-Port. 2008. Non-native speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research* 51: 1480–1493.

Oyama, S. 1976. A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research* 5: 261–283.
Patkowski, M.S. 1980. The sensitive period for the acquisition of syntax in a second language. *Language Learning* 30: 449–472.
Patkowski, M.S. 1990. Age and accent in second language: A reply to James Emil Flege. *Applied Linguistics* 11: 73–89.
Peperkamp, S., and E. Dupoux. 2002. A typological study of stress "deafness". In *Laboratory phonology 7*, ed. C. Gussenhoven and N. Warner, 203–240. Berlin: Mouton de Gruyter.
Peperkamp, S., I. Vendelin, and E. Dupoux. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics* 38(3): 422–430.
Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233–255.
Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. PhD thesis, MIT, Indiana University Linguistics Club.
Pierrehumbert, J. 2002. Word-specific phonetics. In *Laboratory phonology VII*, ed. L. Phonology, 101–139. Berlin: Mouton de Gruyter.
Piper, T., and D. Cansin. 1988. Factors influencing the foreign accent. *The Canadian Modern Language Review* 44: 334–342.
Piske, T., I.R.A. Mackay, and J.E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.
Pisoni, D.B., R.N. Aslin, A.J. Perey, and B.L. Hennessy. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology. Human Perception and Performance* 8: 297–314.
Polka, L. 1995. Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America* 97: 1286–1296.
Polka, L., and J.F. Werker. 1994. Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology. Human Perception and Performance* 20: 421–435.
Singleton, D., and L. Ryan. 2004. *Language acquisition: The age factor*. Clevedon: Multilingual Matters.
Sjölander, K., and J. Beskow. 2000. WaveSurfer – An open source speech tool. In *Proceedings of the 6th International Conference on Spoken Language Processing*. http://www.speech.kth.se/wavesurfer/. Accessed 17 Dec 2009.
So, C.K., and C.T. Best. 2010. Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech* 53: 273–293.
Spaai, G.W.G., and D.J. Hermes. 1993. A visual display for the teaching of intonation. *CALICO Journal* 103: 19–30.
Speech Analyzer. SIL. 2007. [Computer program]. http://www.sil.org/computing/sa/index.htm. Accessed 4 Oct 2009.
Strange, W., and S. Dittmann. 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics* 36: 131–145.
Taniguchi, M., and E. Abberton. 1999. Effect of interactive visual feedback on the improvement of English intonation of Japanese EFL learners. *Speech, Hearing and Language: Work in Progress* 11: 76–89.
Trehub, S.E. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47: 466–472.
Trofimovich, P., and W. Baker. 2006. Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30.
Trofimovich, P., and W. Baker. 2007. Learning prosody and fluency characteristics of L2 speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied PsychoLinguistics* 28: 251–276.

Walley, A.C., and J.E. Flege. 1999. Effect of lexical status on children's and adults' perception of native and non-native sounds. *Journal of Phonetics* 27: 307–332.
Wang, X., and M.J. Munro. 2004. Computer-based training for learning English vowel contrasts. *System* 32: 539–552.
Wayland, R.P., and S.G. Guion. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54(4): 681–712.
Weltens, B., and K. de Bot. 1984. Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech* 271: 79–88.
Werker, J.F., and R.C. Tees. 1983. Developmental change across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology* 37: 278–286.
Werker, J.F., and R.C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development* 7: 49–63.
Werker, J.F., J.H.V. Gilbert, K. Humphrey, and R.C. Tees. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52: 349–355.
Willems, N.J. 1982. *English intonation from a Dutch point of view*. Dordrecht: Foris Publications.
Yu, V.Y., and J.E. Andruski. 2010. A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research* 39(4): 323–344.

Chapter 4
Factors Affecting the Perception and Production of L2 Prosody: Research Results and Their Implications for the Teaching of Foreign Languages

Thorsten Piske

4.1 Introduction

A very large number of studies have shown that second language learners usually speak an L2 with varying degrees of foreign accent (for reviews, see, e.g., Long 1990, Piske et al. 2001, Moyer 2004, Gut 2009). According to Munro (1998, p. 139), a foreign accent can be defined as "non-pathological speech produced by second language […] learners that differs in partially systematic ways from the speech characteristic of native speakers of a given dialect". Similarly, other authors (e.g., Scovel 1969) have used the term foreign accent to refer to the deviations perceived in the pronunciation of non-native speakers from native speaker norms. Moyer (forthcoming) points out that accent is both a collective phenomenon that distinguishes entire speech communities and an individual phenomenon, which reveals a lot about a speaker's personal background and his or her communicative purpose in a certain situation. This is why she defines accent as "a set of dynamic segmental and suprasegmental habits that convey linguistic meaning as well as group, individual and situational identity". According to Moyer (forthcoming), this definition emphasizes that accents serve a linguistic, social and psychological purpose. On the basis of segmental and suprasegmental aspects of speech they deliver semantic content, situate speakers as members of certain groups, affirm identity on a more personal level and establish relative proximity to a speaker's interlocutors in accordance with the situation.

In the present chapter, the focus will be on foreign accent as an individual phenomenon, and in particular, on the ways in which an L2 learner's pronunciation of a second language may correspond to or deviate from the pronunciation norms

T. Piske (✉)
Friedrich-Alexander-Universität, Erlangen-Nürnberg, Lehrstuhl für Fremdsprachendidaktik, Regensburger Str. 160, D-90478 Nürnberg, Germany
e-mail: thorsten.piske@ewf.uni-erlangen.de

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_4,

of native speakers of the L2. Studies examining foreign accent as an individual phenomenon have often been concerned with the question of why some L2 learners manage to speak a second language without a foreign accent, whereas many, if not most, L2 learners retain a foreign accent, even if they have been speaking the L2 for many years (e.g., Purcell and Suter 1980; Flege et al. 1995; Piske et al. 2001; Gut 2009). This question will also be addressed here. First, however, attitudes towards foreign-accented speech and phonetic parameters contributing to the perception of a foreign accent in L2 speech will be explored.

4.2 Attitudes Towards Foreign Accents

As Moyer (forthcoming) points out, an accent reflects a speaker's age and gender, his or her regional, social, cultural and ethnic background and also his or her educational level. On the basis of the foreign accent perceived in the speech of L2 learners, native speakers of the L2 can identify, with relative ease, the L2 learner's first language (L1) background (e.g., Derwing and Munro 1997).

Native speakers of an L2 display different attitudes towards the foreign accents they perceive in non-native speech and thus, also to the speakers of these accents. In more trivial statements, the specific accent perceived in non-native speech is sometimes described as "cute", "funny" or "strange" by native speakers. However, foreign-accented speech may also be stigmatized and have negative effects on an L2 speaker's social status in certain communities. Studies examining attitudes towards the non-native accents of foreign-born teachers have, for example, shown that teachers with foreign accents may be perceived to be less intelligent than teachers without foreign accents by both students and parents (e.g., Nelson 1991). The attitudes native speakers have towards a foreign accent also depend on the particular L1 background that is perceived in the speech of a non-native speaker. A review of studies from the 1970s, 1980s and 1990s by Hamers and Blanc (2000), for example, shows that native speakers of English generally have positive attitudes towards standard native English accents and that they also attribute high rankings for status and intelligence to different Asian and European accents. However, as indicated by the results of a study obtained by Cargile (1997), the attitudes that are expressed towards foreign accents are not only dependent on who is listening to whom, but also on the contexts in which foreign-accented speech occurs. Cargile (1997) examined attitudes towards Mandarin Chinese-accented English in the context of an employment interview and in the context of a college classroom. Cargile (1997) found that a speaker of Chinese-accented English was treated no differently than a speaker of standard American English in the context of an employment interview. In the context of a college classroom, on the other hand, the same Chinese-accented speaker was deemed less attractive than the speaker of standard American English. Cargile also found that Asian American listeners were less evaluatively generous than Anglo-American listeners when they were asked to make estimations of the Chinese-accented speaker's attractiveness.

More negative attitudes towards foreign-accented speech may be due, at least in part, to the prevalent assumption that accent is central to intelligibility and comprehensibility of speech and, in particular, that a strong foreign accent precludes fully intelligible speech. In different studies, Derwing and Munro have examined the relationship between accent, intelligibility and perceived comprehensibility in more detail (e.g., Munro and Derwing 1995a, b, Derwing and Munro 1997). According to the authors, the term 'intelligibility' refers to the extent to which a native speaker understands the intended message. The term 'comprehensibility', on the other hand, "refers to judgments on a rating scale of how difficult or easy an utterance is to understand" (Derwing and Munro 1997, p. 2). In two of their studies (Munro and Derwing 1995a, and Derwing and Munro 1997), they found that accent ratings assigned by native speakers of English to speech samples produced by non-native speakers of English from different L1 backgrounds were harsher than perceived comprehensibility ratings, which in turn were harsher than intelligibility scores. Moreover, the results of correlational analyses examining the relationship between foreign accent, perceived comprehensibility and intelligibility led the authors to conclude that these three dimensions are closely related but not equivalent and that a strong degree of L2 foreign accent does not necessarily interfere with intelligibility. Nevertheless, some foreign-accented but fully intelligible utterances may require more effort or processing time on the listeners' side, which may lead native speakers to rate these utterances as difficult to understand (e.g., Munro and Derwing 1995b; Derwing and Munro 1997).

The results and observations discussed so far indicate that the attitudes native speakers express towards a foreign accent may be influenced by factors such as the L1 background they perceive in the speech of a non-native speaker, the context in which foreign-accented speech occurs and the effort it may require to understand foreign-accented speech. One question that has not been addressed yet is how non-native speakers themselves may feel about their own pronunciation of a second language. According to Flege (1987), foreign-accented speech may serve quite different social functions for child and adult L2 learners. He points out that social factors, along with several other factors, may help to explain why in most studies child L2 learners have been found to be more successful than adult L2 learners in the pronunciation of an L2. According to Flege (1987), child L2 learners may feel stronger social pressure from their peers than adult L2 learners to pronounce an L2 in an authentic way because children may feel a greater need than adults to participate fully in the culture associated with the second language. Moreover, children may be more willing than adults to communicate with native speakers of an L2 because they are less fearful than adults of making mistakes or of being ridiculed by native speakers for using the L2 in an ineffective way (see also Schumann 1975).

Results obtained by Butler (2007) suggest that children learning an L2 also prefer to have L2 teachers who speak the L2 without a foreign accent. Butler (2007) used a matched-guise technique to examine Korean elementary school students' attitudes towards American-accented English and Korean-accented English. She found that the children learning English in two elementary schools in Korea thought that the American-accented English guise had a better pronunciation of English, was

relatively more confident in her use of English, would focus more on fluency than on accuracy and would use Korean less often in the English class. Moreover, the Korean elementary school students also expressed that they preferred to have the American-accented English guise as their English teacher. On the whole then, the results and assumptions just reported suggest that child L2 learners may be more willing than adult L2 learners to communicate with native speakers of the L2 and that this might be one of the reasons why child L2 learners have usually been found to be more successful than adult L2 learners in developing an accurate pronunciation of the L2.

According to Flege (1987), adult L2 learners, in contrast to children, may even have good reasons to maintain a foreign accent. Maintaining a foreign accent could, for example, help adult L2 learners to preserve their ethnic and cultural identity in a predominantly L2-speaking environment. Moreover, social blunders might be tolerated more readily in an obvious adult foreigner who speaks an L2 with a strong degree of foreign accent than in a foreigner whose L2 speech is characterized by no or just a slight foreign accent. Of course, these assumptions should not be taken to mean that adults might not be highly motivated to learn an L2 well. Both Bongaerts et al. (1997) and Moyer (1999), for example, examined late L2 learners, most of whom had worked as teachers of the L2 and considered it necessary to speak the L2 without a foreign accent. Moyer (1999) found a strong correlation between the variable "professional motivation" and degree of L2 foreign accent. However, only one of the 24 participants of her study was found to speak the L2 without a foreign accent. In the study by Bongaerts et al. (1997), 5 out of 11 highly motivated late L2 learners received foreign accent ratings comparable to those obtained for a control group of native speakers of the L2. Results like these show that a high level of motivation may have an influence on learning to pronounce an L2 well, but that it does not automatically lead to accent-free L2 speech (see also Gut 2009).

In summary, both native and non-native speakers display varying attitudes towards foreign-accented speech. The attitudes of native speakers towards the foreign accent they perceive in the speech of L2 learners appears to be dependent on factors such as the L1 background of the L2 speaker, the contexts in which foreign-accented speech is encountered and the effort required to understand the speech of L2 learners. The attitudes non-native speakers themselves have towards their own pronunciation of an L2 appear to depend very much on factors such as "integrative" or "professional motivation". Parameters contributing to the perception of a foreign accent in non-native speech will be explored in the next part of this chapter.

4.3 Phonetic Parameters Contributing to the Perception of a Foreign Accent

In many studies examining the phenomenon of perceived foreign accent in the speech of L2 learners, native speakers of the L2 have been recruited to use a rating scale in order to "measure" the degree of foreign accent they perceive in samples of

non-native speech. When asked about the phonetic parameters contributing to the perception of a foreign accent, native-speaking raters usually report that they perceive non-native speech as foreign-accented because of both segmental as well as suprasegmental/prosodic errors produced by L2 learners. In addition, the degree of foreign accent perceived in nonnative speech is also likely to be determined by the fluency with which bilinguals produce a second language, i.e., by pause and hesitation phenomena, such as silent and filled pauses, repetitions, false starts and rate of speech (see, e.g., Hieke 1980; Derwing et al. 2009). Flege (1984) reported that native English listeners were able to identify a French accent in only one syllable, i.e., [ti] or [tu]. According to Flege (1984), the French accent may have been revealed by subcategorical phonetic differences or by differences in vowel duration.

Several studies have taken a closer look at the relative contribution of segmental parameters, prosodic parameters and fluency to degree of foreign accent in L2 speech. Derwing and Munro (1997), for example, asked 26 native English listeners from Canada to use 9-point rating scales to rate picture narrations produced by 48 L2 learners of English from four different L1 backgrounds (Cantonese, Japanese, Polish and Spanish) for degree of L2 foreign accent and degree of comprehensibility. 13 listeners were also asked to answer open-ended questions regarding which factors they considered important when judging degree of L2 foreign accent. Whereas 12 of the 13 listeners believed that segmental deviations strongly contribute to the perception of a foreign accent in non-native speech only 3 of the 13 listeners thought that prosodic features strongly contribute to the perception of a foreign accent.

The assumption that segmental deviations are the primary source of foreign accent is supported by the results of a study by Jilka (e.g., 2000, 2002), who specifically examined the contribution of prosodic aspects of speech to the perception of a foreign accent. The results Jilka (2000, 2002) obtained in analyses of foreign-accented data collected from 10 native speakers of German learning American English as an L2 and 10 native speakers of American English learning German as an L2 led him to conclude that prosody significantly contributes to the perception of a foreign accent and that intonation is the most important prosodic factor in the perception of foreign accent in non-native speech. At the same time, the foreign accent ratings Jilka obtained for two types of synthesized stimuli, which were meant to represent either only a segmental foreign accent or only an intonational foreign accent, suggested that segmental aspects of foreign accent are perceived more strongly by native speakers of an L2 than intonational ones.

A number of studies have examined to what extent temporal aspects of oral production contribute to the perception of a foreign accent in non-native speech. In their study examining the relationship between accent, intelligibility and comprehensibility, Derwing and Munro (1997) found that speaking rate was negatively correlated with both degree of L2 foreign accent and degree of comprehensibility. According to Derwing et al. (2009, p. 534), fluency is "one of the most noticeable dimensions along which second language learners differ". The degree of fluency or dysfluency that is perceived in the speech of both native and non-native speakers is primarily determined by the number of pauses, hesitation phenomena and self-repetitions occurring in their speech and by the speaking rate at which they produce

certain speech materials. Several studies have found that the overall speaking rate is slower in non-native speech than it is in native speech. Munro and Derwing (1995b), for example, compared the duration of English speech materials produced by native speakers of Mandarin Chinese and by native speakers of English and they found that the native Mandarin Chinese speakers' utterances were longer than those produced by the native speakers of English. Similarly, Guion et al. (2000) measured the duration of sentences produced by Italian-English and Korean-English bilinguals who had already been living in an English-speaking environment for 12–35 years at the time of testing. They found that for both the native speakers of Italian and the native speakers of Korean, the duration of English sentences was positively correlated with their age upon arrival in an English-speaking country. In other words, English sentences produced by native speakers of English were significantly shorter than English sentences produced by native speakers of Italian and native speakers of Korean who had started to learn English rather late in life, i.e., in adolescence or early adulthood.

Measurable duration differences between English phrases and sentences produced by native English children and adults and by native Japanese children and adults were also reported by Aoyama and Guion (2007). They found that the absolute duration of syllables and whole utterances tended to be longer in native Japanese speakers' English utterances than in native English speakers' utterances and that the durations of function words were proportionally longer in native Japanese speakers' utterances than in native English speakers' utterances. Aoyama and Guion (2007) also reported that children's utterances were longer in general than adults' utterances and syllables contrasted less clearly within an utterance in child speech compared to adult speech. According to the authors, the observed differences between child and adult speech indicate that there may be developmental changes in the production of prosodic aspects of speech.

In a study examining 14 L2 learners of German from different L1 backgrounds (English, Italian, Romanian, French and Chinese), Gut (2003) found that speakers whose L1 does not mark syllables as reduced or short produce such syllables significantly less often in their L2. According to Gut (2003), the results of her study suggest that native language influence on L2 speech is not only reflected in the production of L2 consonants and vowels but also in the production of L2 prosody. In addition, results obtained for more and less advanced L2 learners of German led Gut (2003, p. 149) to conclude that "low proficiency precludes near-native production of prosodic features".

In a study examining both L2 learners of German and English, Gut (2009) examined the relationship between different linguistic domains of non-native speech. She found little correlation between different areas of non-native phonology such as consonantal processes, vowel reduction and speech rhythm and intonation. According to Gut (2009), the results of her study support the earlier finding by Wennerstrom (1998) and Lléo and Vogel (2004) that non-native speakers can pronounce native-like features in one area of phonology although they may show a high error rate in another. Gut (2009) also examined the relationship between non-native speakers' abilities in phonological and non-phonological areas and found that some

phonological and non-phonological features were indeed correlated, but only in retellings, not in free speech. On the whole, her findings led Gut (2009) to draw the conclusion that just like near-native abilities in one linguistic area cannot predict near-native abilities in another linguistic area, a high rate of errors in one area of language is not necessarily correlated with an equally high rate of errors in another linguistic area.

Finally, Derwing et al. (2009) examined whether speakers who are extremely fluent in their L1 would also be more fluent than their peers in the L2. In their study, they examined L1 fluency and L2 English fluency at three times over 2 years in native Mandarin- as well as native Russian- and native Ukrainian-speaking adult immigrants to Canada. Fluency ratings obtained from trained judges of picture narrations produced by the L2 speakers of English indicated a relationship between L1 and L2 fluency in the initial stages of L2 exposure. More specifically, pauses per second, speech rate and pruned syllables per second, but not vowel durations, were all related to the listeners' judgments in both languages. However, at later stages, the correlations between the judgments was lower and not statistically significant. According to Derwing et al. (2009, p. 554), the results of their study suggest "that a straightforward relationship between L1 and L2 fluency cannot be expected". Apparently. several factors, such as the structural properties of the L1 itself, the proficiency levels of the speakers, the amount and type of exposure to the L2 and cognitive factors all contribute to fluency development in an L2.

In summary, according to the results reported here, segmental parameters, prosodic parameters and fluency phenomena all contribute to the perception of a foreign accent in L2 speech. The results of some studies suggest that segmental deviations contribute more strongly to the perception of a foreign accent in non-native speech than prosodic deviations, and just like segmental deviations, prosodic deviations appear to be influenced by the phonological structure of a learner's native language. As regards the role of fluency in the perception of a foreign accent in the speech of L2 learners, there is apparently no straightforward relationship between L1 and L2 fluency. Apart from the structural properties of the learners' native languages, several factors such as the proficiency levels of L2 speakers, amount and type of exposure to the L2 and different cognitive factors all appear to contribute to both fluency development and to the development of L2 speech production and perception in general. The next section will take a closer look at studies that have examined the influence of different factors on L2 speech learning.

4.4 Difficulties in Interpreting the Findings of Foreign Accent Research

Since the end of the 1960s, a large number of increasingly detailed experimental studies have examined various subject and phonetic variables that have been claimed to affect the degree of foreign accent with which non-native speakers pronounce an L2. Many of these studies have mainly focused on the examination of subject

or speaker variables (for overviews, see Piske et al. 2001, Gut 2009 and different chapters in Hansen Edwards and Zampini 2008). The one subject variable that has received the most attention in previous research is age of first extensive exposure to the L2, which has often been indexed by the participants' age of arrival (AOA) in an L2-speaking country. Other subject variables that have also been examined in a relatively large number of L2 studies include the participants' chronological age, their length of residence in a predominantly L2-speaking community/country, their gender, their L1 background, amount of formal instruction, amount and type of training non-native speakers have received in the perception and production of segmental and prosodic aspects of L2 speech, amount of L1 and L2 use, language learning aptitude and motivational and affective variables. Subject variables whose role in L2 speech has been investigated in only a smaller number of studies include L1 ability or proficiency, field independence, right hemispheric specialization and introversion/extroversion.

Some authors have emphasized that the results obtained in studies of L2 speech perception and production are not only determined by all the subject variables just listed but they are also dependent on the whole context of an experiment, that is, they are also dependent on contextual or "speaker-independent" variables (see, e.g., Munro and Derwing 1998; Levi et al. 2007). These authors have, for example, looked at the way in which the results of certain experiments were dependent on the presence or absence of orthographic input during the presentation of stimuli, the general frequency of the words participants had to perceive or produce, the speaking rate at which certain speech materials were produced or whether participants happened to be in what Grosjean (e.g., 2001) has called a 'monolingual' or a 'bilingual mode' at the time of testing.

Finally, a very large number of studies have also examined the role of phonetic factors in second language speech learning. These studies have typically examined how L1 and L2 vowels, L1 and L2 consonants or L1 and L2 prosodic parameters are related to each other and how L2 speech learning is affected by these phonetic relationships (see, e.g., Gut 2009 and different chapters in Bohn and Munro 2007).

According to the results of studies examining immigrant L2 learners, age of L2 learning appears to be the variable that plays the most important role in the perception and production of L2 speech. However, the relative importance of other variables is uncertain. This is particularly true for many of the subject variables mentioned above. Why is it still quite difficult to draw any stronger conclusions about the influence that several subject variables have on overall degree of L2 foreign accent? The following paragraphs discuss three types of reasons that make it difficult to interpret the findings of previous L2 foreign accent studies.

First, as pointed out by Piske et al. (2001), direct comparisons across studies are often problematic because the studies of L2 foreign accent that have been carried out to date have differed greatly in terms of design and methodology. The subject populations studied in previous research have, for example, differed in terms of the target L2 they had learned, in terms of their native languages, in terms of the amount and type of experience they had with the target L2 and also in terms of their degree of motivation to learn to speak an L2 well. Moreover, in studies in which native

speakers of an L2 were asked to evaluate speech samples produced by non-native speakers, different rating techniques have been used. In many studies equal-appearing interval scales have, for example, been employed in order to measure the degree of foreign accent that can be perceived in speech produced by L2 learners. The gradients on such scales have varied greatly though, ranging from 3-point to 10-point scales, and in studies using sliding scales, values between 0 and 255 were even returned. Not all studies used equal appearing interval scales. Some adopted direct magnitude estimation. Moreover, in some studies, native listeners were recruited to evaluate non-native speech, in other studies, this task was completed by expert raters, such as ESL teachers or linguists.

As also pointed out by Piske et al. (2001), studies of L2 foreign accent have also differed greatly in terms of the techniques used to elicit non-native speech. Each of these techniques involves specific problems. In most studies, the participants have been asked to read either sentences, paragraphs or individual words. The use of read speech in L2 speech production research is problematic, however, because just like native speakers, non-native speakers may differ in reading ability and this is why read speech may not always reveal how accurately non-native speakers can pronounce an L2 when they speak it more or less spontaneously. In some studies (e.g., Oyama 1976; Thompson 1991), read speech was indeed judged to be more strongly foreign-accented than conversational speech samples. Sometimes imitation techniques have also been used to elicit non-native speech. Some studies have used a direct imitation technique. In several other studies, a delayed repetition technique has been used. When repetition techniques are used, the participants usually have to repeat speech materials after listening to a native speaker model. What remains uncertain in these studies, of course, is how accurately the participants would have pronounced certain speech materials if these materials had not been modelled beforehand by native speakers.

Most researchers will probably acknowledge that conversational speech, that is, speech produced more or less spontaneously, should represent the most important criterion for success in learning to pronounce and perceive a second language. Unfortunately, however, the number of studies in which the participants have been asked to listen to or produce conversational speech is relatively small. This is probably not surprising because it is difficult to analyse conversational speech under controlled conditions. Piske et al. (1999, 2011) developed a technique to examine L2 vowel production under controlled conditions. The participants of this study were native Italian learners of English, who had all been living in Canada for a minimum of 18 years at the time of testing but who differed in terms of age of arrival in Canada and in terms of self-reported amount of L1 use. They were asked to produce English vowels in conversational speech and in formal experiments in which they had to produce real English words and non-words elicited with the help of words that were modelled aurally and that also appeared on a written list. One of the most interesting results of the study was that the accuracy with which English vowels such as /ɪ/, /ʊ/ and /ə˞/, which do not occur in Italian, were produced by a group of early L2 learners who still spoke their L1 Italian frequently, depended on the context in which the production of vowels had been elicited. It was found that this

group of early L2 learners was able to produce vowels such as /ɪ/, /ʊ/ and /ɚ/ accurately in real words and in conversational speech but not in non-words. Such findings suggest that results obtained under highly controlled experimental conditions, in which the production of L2 speech is elicited with the help of speech materials presented orally and/or in written form, may not always reveal how accurately L2 learners can pronounce an L2 in what could be called "real world conditions", i.e., if they are asked to talk about something more or less spontaneously.

Given the large differences between studies in terms of design and methodology, it is not surprising that previous studies examining factors affecting L2 speech perception and production have at times produced results that appear to conflict and such conflicting results are, of course, not easy to interpret. A second reason why it is sometimes difficult to draw any stronger conclusions regarding the relative contribution of different subject variables to degree of L2 foreign accent is that many of the subject variables that have been examined in previous research tend to be confounded. For example, age of arrival, which appears to be one of the strongest predictors of success in L2 speech learning, has repeatedly been found to be confounded with chronological age, length of residence in an L2-speaking country, amount of L1 and L2 use and amount of instruction received in the L2-speaking country (e.g., Flege 1998, 2009; Flege et al. 1999; Piske et al. 2001). A technique that can be used to control for expected confounds between one subject variable and other variables is the 'matched subgroup technique', which is a statistical method of refining group results by analysing smaller groups within a large group, which differ for one specified variable (e.g., age of arrival) but which match for variables confounded with that variable (e.g., Flege et al. 1999; Flege 2009). Unfortunately, however, techniques like the matched subgroup technique have only been used in a relatively small number of studies examining age effects on L2 learning, so it remains uncertain whether it was really age or variables confounded with age that determined the participants' performance in these studies.

One of the reasons why age variables have often been examined in previous L2 foreign accent studies is probably that they are easy to measure. Other variables, on the other hand, are difficult to measure precisely and this is the third reason why it is sometimes difficult to draw any stronger conclusions regarding the relative contribution of certain subject variables on degree of L2 foreign accent. Variables that are difficult to measure include motivational and affective variables, variables that relate to a special language learning aptitude, language use variables and quantity and quality of L2 input.

If one considers how learner input has been assessed in previous research, it becomes obvious how difficult it is to measure certain subject variables precisely. In several studies, Flege and his colleagues have tried to assess the quantity and quality of the L2 input L2 learners had been exposed to by examining two variables which appear to be at least indirectly related to L2 input. These two variables are length of residence in an L2-speaking country and amount of L1 and L2 use. One might think that amount of L2 input should be correlated with length of residence in an L2 speaking country because the longer an L2 learner has lived in an L2-speaking country, the more L2 input this person could probably be expected to have received.

However, as discussed by Piske et al. (2001), not every study has shown a significant effect of length of residence on degree of L2 foreign accent and in those studies where an effect of length of residence was found, the effects of this variable tended to be small. The results of several studies (e.g., Winitz et al. 1995; Flege and Liu 2001; Flege et al. 2006; see also Flege 2009) suggest that larger effects of length of residence on degree of L2 foreign accent will be obtained only for immigrant L2 learners who receive a substantial amount of native-speaker input during their stay in an L2-speaking country.

It also appears reasonable to assume that amount of L2 input should also be correlated with language use variables. The more an L2 learner uses an L2, the more this person may be expected to receive input from other speakers of the L2. Different studies (e.g., Flege et al. 1995; Piske et al. 2001) have shown that language use variables do indeed affect degree of L2 foreign accent, but that amount of L1 or L2 use is a less important predictor of L2 speech learning than age of L2 learning. Although it has usually been found that age of L2 learning has stronger effects on degree of L2 foreign accent than language use variables, it would, according to Flege (2009), be imprudent to conclude that amount of L1 or L2 use are unimportant for L2 speech learning. This is because so far amount of L1 or L2 use has to our knowledge never been measured directly, but the language use estimates reported in different studies were all based on immigrants' self-reports, which may, of course, be subject to error. Moreover, if immigrants report that they frequently use an L2, this does not automatically mean that they frequently communicate with native speakers of the L2. It is possible that they mainly use the L2 when communicating with other non-native speakers, and due to the frequent exposure to accented L2 speech, they may develop norms for the segmental and prosodic parameters of the L2 that differ from the norms of native speakers. Taken together, all of this suggests that variables such as length of residence and amount of L1 and L2 use only provide a very rough estimate of the amount of L2 input learners may have received from native speakers. What the findings just reported do show is that in order to learn to accurately pronounce an L2 it is not that important how long an immigrant lives in a country where the L2 is spoken, it is more important with whom the immigrant spends a lot of time communicating in the L2 while living in that country.

As has been described in this chapter, there are three types of reasons why it is still quite difficult to draw any stronger conclusions regarding the relative contribution of different variables and in particular different subject variables, to degree of L2 foreign accent. Nevertheless, it is important to note that some subject variables have relatively often been found to have a significant influence on L2 speech learning, whereas other variables have only rarely been identified as significant predictors of success in learning to produce and perceive an L2. Variables that have repeatedly been found to significantly affect L2 speech learning include a) age of first extensive exposure to the L2, which has often been indexed by immigrants' age of arrival in an L2-speaking country, b) L1 background, i.e., the phonetic/phonological characteristics of learners' first or native language(s), c) training in the perception and production of L2 speech and d) variables that are probably associated with quantity and quality of L2 input, such as length of residence and amount of L1

Table 4.1 Characteristics of the ten participants who received the best foreign accent ratings on a 9-point rating scale (1 = very strong foreign accent; 9 = no foreign accent) in a study by Piske et al. (2001)[a]

AveFa	Group	AOA	L1 Use	LOR
8.74	Early-high	1.5	67	36
8.48	Early-high	5	30	39
8.48	Native English	–	–	–
8.48	Native English	–	–	–
8.37	Native English	–	–	–
8.33	Native English	–	–	–
8.30	Early-low	12	12	39
8.30	Native English	–	–	–
8.26	Early-low	2	1	40
8.22	Early-low	4.5	6	46

AveFa = Average foreign accent rating given to participant, Group = Group participant was assigned to based on her/his AOA and amount of self-reported L1 use, AOA = age of arrival in Canada, in years, L1 Use = self-reported percentage use of Italian, LOR = length of residence in Canada, in years, Early high = Relatively early AOA (group mean = 8 years), relatively high percentage of L1 use (group mean = 43%), Early low = relatively early AOA (mean = 7 years), relatively low percentage of L1 use (mean = 7%).

[a]The range of ratings given to the 90 participants examined in the Piske et al. (2001) study indicates how much individual subjects in each of the five groups of participants differed from each other in terms of the degree of foreign accent with which they spoke English. The average foreign accent ratings given to the 18 subjects in the Native English group ranged from 7.11 to 8.48, those given to the 18 subjects in the Early-low group ranged from 4.33 to 8.3 and those given to the 18 subjects in the Early-high group ranged from 3.11 to 8.74. In addition, there were two groups of late learners (mean AOA = 20 years) also differing in terms of self-reported percentage use of Italian (Late-low mean = 10% vs. Late-high mean = 53%). The average foreign accent ratings given to 18 subjects in the Late-low group ranged from 1.81 to 6.67 and those given to 18 subjects in the Late-high group ranged from 1.44 to 5.56.

and L2 use. It is important to note, however, that according to the results obtained for individual learners, each of these variables may be of different importance for different individuals. In a study examining four groups of native Italian immigrants to Canada (18 per group), Piske et al. (2001), for example, found that age of L2 learning and amount of L1 use had significant, independent effects on degree of L2 foreign accent. As far as the effects of amount of L1 use are concerned, the results of their study suggested that frequent use of the L1 correlated with a relatively strong degree of L2 foreign accent. Surprisingly, however, a closer look at the results obtained for individual learners participating in the Piske et al. (2001) study reveals that two participants who received the best foreign accent ratings in the study belonged to a group of non-native speakers who had moved from Italy to Canada in early childhood (i.e., at ages 1.5 and 5.0 years, respectively) but who continued to use their L1 Italian relatively often (i.e., 67% and 30% of the time; see Table 4.1). This means that these two subjects, one of whom received even better foreign accent ratings than any of the subjects in a control group of 18 native speakers of English, did not support the more general finding of the Piske et al. (2001) study that continued frequent use of the L1 has negative effects on degree of L2 foreign accent.

One of these two participants worked as a teacher in Canada and she was, therefore, probably highly motivated to pronounce her L2 English well. Such biographical details about participants' profession, hobbies, etc., are usually not reported in studies comparing groups of learners but they may help to explain why individual learners are highly successful in learning to pronounce an L2 accurately, although these learners do not show all the characteristics that would be expected of highly successful learners.

4.5 Implications for Foreign Language Teaching

As Piske (2008) points out, the learning conditions found in most foreign language classrooms make it quite difficult for learners to learn to pronounce an L2 well. Foreign language students usually hear their classmates' and often also their teachers' incorrect pronunciations and are, therefore, likely to develop norms for the pronunciation of the L2 that differ from native speakers' norms. Even if learners begin to learn a foreign language in preschool or elementary school, their perceptual abilities have, according to the results of studies examining infant speech perception (for a review, see e.g., Burnham et al. 2002), already been modified in the direction of those sound contrasts that are used for distinctive purposes in their L1. If foreign language learners have already learned to read and write in their native language, they may pronounce certain graphemes occurring in L2 words as they are pronounced in their L1 (e.g., Bassetti 2009). Moreover, the use of phonetic symbols to raise students' awareness of similarities and differences between their L1 and L2 sound systems may also contribute to non-target-like pronunciations because phonetic symbols may insufficiently represent acoustic similarities and differences between L1 and L2 sounds (e.g., Steinlen 2005). However, as Piske (2008) also points out, based on the findings of research examining factors affecting degree of L2 foreign accent, it is possible to draw several conclusions regarding learning conditions that should help foreign language learners to develop an accurate L2 pronunciation. Among other things, these findings, for example, suggest that foreign language students should be given the opportunity to learn a second language early in life, that they should receive a substantial amount of native-speaker input and that they should be encouraged to use the new language as often as possible. Moreover, the finding that the phonological structure of a learner's native language has an influence on the pronunciation of both segmental as well as prosodic aspects of L2 speech indicates that the common practice of providing all the students in a foreign language classroom with the same type of training in the perception and production of L2 speech is problematic. If there are students from different L1 backgrounds in a foreign language classroom, they should rather receive different types of training depending on the specific differences between the phonological structures of their L1 and the L2.

In addition, the results of many L2 foreign accent studies suggest that training in the production and perception of L2 speech can also have positive effects on

the pronunciation of an L2. Neufeld (1978), for example, reported that intensive perceptual training had a positive influence on native English learners' production of articulatory and prosodic features of three non-Indo-European languages. In a study examining grade 3 and 4 English-as-a-second-language learners, Trofimovich et al. (2008) found that after a period of 2 years, long-term comprehension practice in listening and reading had positive effects on L2 fluency and comprehensibility. Gut (2009) analysed the effects of a training course consisting of the parts 'theory', 'perception' and 'production' on L2 speech learning. On the one hand, she found that foreign accent ratings that were given to relatively advanced L2 learners and that were recorded before and after the course did not improve, which suggested that the course appeared not to have been successful. On the other hand, she also found that the more advanced learners examined in her study profited from theoretical training because they knew significantly more about L2 prosody after they had taken the course than before taking the course. Hardison (2005) found that native Chinese learners of English showed significant improvement in the use of English prosody in novel natural discourse when they received focused prosody training, using selected samples from the learners' own oral L2 productions. The results of different studies (e.g., Derwing et al. 1998, Missaglia 1999; Derwing and Rossiter 2003; Hahn 2004) suggest that learners profit to a larger extent from instruction that focuses on suprasegmental rather than segmental aspects of pronunciation. Missaglia (1999, see also Missaglia 2007), for example, found that prosody centred training based on the Contrastive Prosody Method had an ameliorative effect on both prosodic and segmental aspects of native Italian learners' pronunciation of German. Moreover, the results of studies by Derwing et al. (1998) and Derwing and Rossiter (2003) examining how intelligible the speech of L2 learners was judged to be showed that L2 learners who received instruction focusing on stress and rhythm were judged to be easier to understand than L2 learners who received instruction that focused on individual sounds. Although the L2 learners who received instruction focusing on individual sounds were found to produce these sounds more accurately, this did not seem to increase the intelligibility of their L2 speech.

On the whole, the results of studies examining the effects of training on L2 speech development suggest that training, and in particular training focusing on prosodic aspects of speech, may have positive effects on the degree of foreign accent with which learners pronounce an L2. However, as Yule and MacDonald (1994, pp. 116–117) point out, the results of several studies also suggest that learners may show rather individual reactions to different types of pronunciation training so that "the learner may represent a more powerful variable in such studies than the type of teaching method involved".

Finally, based on a review of studies in which learners were asked which type of pronunciation training they found useful, Settinieri (2008) concluded that L2 learners prefer training which involves the following characteristics: individual attention, focus on segmental and suprasegmental features, language and language learning awareness raising, meaningful and authentic exercises and visual support.

4.6 Summary and Conclusions

Non-native speakers usually pronounce second languages with varying degrees of L2 foreign accent. According to the findings of the studies reported here, segmental parameters, prosodic parameters and different phenomena related to fluency all contribute to the perception of a foreign accent in non-native speech. The results of some studies indicate that segmental deviations from native speaker norms contribute more strongly to the perception of a foreign accent in L2 speech than do prosodic deviations. However, on the basis of the available empirical evidence it is very difficult to quantify the relative contribution of segmental and prosodic parameters to degree of L2 foreign accent.

The findings of several studies suggest that the degree of L2 foreign accent with which learners pronounce an L2 is influenced by factors such as age of first intensive exposure to the L2, L1 background, training in the perception and production of L2 speech and by variables that are probably associated with quantity and quality of L2 input, such as length of residence in an L2-speaking community and relative amount of L1 and L2 use. If and to what extent other factors contribute to degree of L2 foreign accent is difficult to determine for different reasons, which have been discussed in this paper. Moreover, an examination of the results obtained for individual learners in studies of L2 foreign accent reveals that each of the factors that have been found to affect L2 pronunciation accuracy may be of different importance for different individuals.

Both native and non-native speakers of an L2 may show rather positive or rather negative attitudes towards foreign-accented speech. The attitudes native speakers display towards the foreign accent they perceive in the speech of an L2 learner appear to depend on the L1 background of the learner, on the context in which they encounter the learner and on the degree of effort required to understand her or his L2 speech. The attitudes non-native speakers have towards their own pronunciation of an L2 appear to be largely dependent on motivational variables and, in particular, on whether learners consider it important to learn to pronounce an L2 well because of job-related reasons and/or because they want to integrate into an L2-speaking community.

The observation that the attitudes people display towards L2 learners may depend on the foreign accent they perceive in the speech of these learners indicates that it is important not to ignore L2 pronunciation accuracy in foreign-language classrooms. The results of studies examining factors affecting degree of L2 foreign language suggest that at school, students should be given the opportunity to learn a second language early in life, that they should receive a substantial amount of native speaker input, that they should be encouraged to use the new language as often as possible and that they should receive training in the production and perception of L2 speech that takes into account the specific differences between the phonological structures of the L2 and the different L1s that may be found in the same classroom. Finally, the results of a few studies also suggest that learners profit more from instruction that focuses on suprasegmental rather than segmental aspects of pronunciation. However, just like a certain factor that has been found to affect L2 speech learning may not be

of equal importance for all L2 learners, a specific type of training that has been found to have positive effects on L2 pronunciation accuracy may not be equally suitable for all students in a foreign language classroom.

It is difficult to speculate about the pragmatic implications of the research reported here. On the basis of previous foreign accent research it is, for example, impossible to determine to what extent the degree of foreign accent with which non-native speakers pronounce an L2 may hinder them in expressing their intentions in certain contexts. In order to obtain more concrete findings about the relationship between foreign accents and their possible influence on the pragmatic competences of non-native speakers or about foreign accents and their possible influence on listeners' interpretations of utterances characterized by a foreign accent, experiments are required in which native listeners are asked to interpret utterances produced in specific contexts by non-native speakers differing in the degree of foreign accent with which they speak an L2. Such experiments could also examine to what extent non-native speakers from different L1 backgrounds transfer "prosodic strategies" used in order to express certain intentions in the L1 when using the L2. And, of course, the effects of different types of training could also be examined in such experiments. In order to increase their pragmatic competences in an L2, learners should receive training in which they are asked to adjust their production of a certain utterance depending on the context in which this utterance occurs and training in which they are made aware of the ways in which native speakers adjust the production of an utterance depending on the situation in which the utterance is made. In most foreign language classrooms, the role different contexts may play in the production and perception of L2 speech does not seem to be addressed in any more detail when students receive training in the production and perception of L2 speech.

References

Ayoama, Katsura, and Susan G. Guion. 2007. Prosody in second language acquisition. Acoustic analyses of duration and F0 range. In *Language experience in second language speech learning. In honor of James Emil Flege*, ed. Ocke-Schwen Bohn and Murray J. Munro, 281–297. Amsterdam: John Benjamins.

Bassetti, Benedetta. 2009. Orthographic input and second language phonology. In *Input matters in SLA*, ed. Thorsten Piske and Martha Young-Scholten, 191–206. Bristol: Multilingual Matters.

Bohn, Ocke-Schwen, and Murray J. Munro, eds. 2007. *Language experience in second language speech learning. In honor of James Emil Flege*. Amsterdam: John Benjamins.

Bongaerts, Theo, Chantal van Summeren, Brigitte Planken, and Erik Schils. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–465.

Burnham, Denis, Michael Tyler, and Stephanie Horlyck. 2002. Periods of speech perception development and their vestiges in adulthood. In *An integrated view of language development: Papers in honor of Henning Wode*, ed. Petra Burmeister, Thorsten Piske, and Andreas Rohde, 281–300. Trier: WVT – Wissenschaftlicher Verlag Trier.

Butler, Yuko G. 2007. How are non-native-English-speaking teachers perceived by young learners? *TESOL Quarterly* 41: 731–755.

Cargile, Aaron C. 1997. Attitudes towards Chinese-accented speech. An investigation in two contexts. *Journal of Language and Social Psychology* 16: 434–443.

Derwing, Tracey M., and Murray J. Munro. 1997. Accent, intelligibility and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition* 19: 1–16.

Derwing, Tracey M., and Marian J. Rossiter. 2003. The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning* 13: 1–17.

Derwing, Tracey M., Murray J. Munro, and Grace Wiebe. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48: 393–410.

Derwing, Tracey M., Murray J. Munro, Ronald I. Thomson, and Marian J. Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31: 533–557.

Flege, James E. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76: 692–707.

Flege, James E. 1987. A critical period for learning to pronounce foreign languages? *Applied Linguistics* 8: 162–177.

Flege, James E. 1998. The role of subject and phonetic variables in second language learning. In *Papers from the 34th annual meeting of the Chicago Linguistic Society,* Vol. II, *The panels*, ed. M. Catherine Gruber, Derrick Higgins, Kenneth Olson, and Tamra Wysocki, 213–232. Chicago: Chicago Linguistic Society.

Flege, James E. 2009. Give input a chance! In *Input matters in SLA*, ed. Thorsten Piske and Martha Young-Scholten, 175–190. Bristol: Multilingual Matters.

Flege, James E., and Serena Liu. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition* 23: 527–552.

Flege, James E., Murray J. Munro, and Ian R.A. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97: 3125–3134.

Flege, James E., Grace H. Yeni-Komshian, and Hua Liu. 1999. Age constraints on second language acquisition. *Journal of Memory and Language* 41: 78–104.

Flege, James E., David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and Kimiko Tsukada. 2006. Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics* 34: 153–175.

Grosjean, François. 2001. The bilingual's language modes. In *One mind, two languages: Bilingual language processing*, ed. Janet L. Nicol, 1–22. Oxford: Blackwell.

Guion, Susan G., James E. Flege, Serena H. Liu, and Grace H. Yeni-Komshian. 2000. Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics* 21: 205–228.

Gut, Ulrike. 2003. Prosody in second language speech production: The role of the native language. *Fremdsprachen lehren und lernen* 32: 132–152.

Gut, Ulrike. 2009. *Non-native speech. A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt am Main: Peter Lang.

Hahn, Laura D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38: 201–223.

Hamers, Josiane F., and Michel H.A. Blanc. 2000. *Bilinguality and bilingualism*, 2nd ed. Cambridge: Cambridge University Press.

Hansen Edwards, Jette G., and Mary L. Zampini, eds. 2008. *Phonology and second language acquisition*. Amsterdam: John Benjamins.

Hardison, Debra M. 2005. Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal* 22: 175–190.

Hieke, Adolf E. 1980. Aspects of native and non-native fluency skills. PhD dissertation, University of Kansas.

Jilka, Matthias. 2000. The contribution of intonation to the perception of a foreign accent, vol. 6, no. 3. University of Stuttgart: Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (AIMS).

Jilka, Matthias. 2002. Testing the contribution of prosody to the perception of a foreign accent. In *New Sounds 2000. Proceedings of the fourth international symposium on second-language speech*, ed. Allan James and Jonathan Leather, 199–207, University of Klagenfurt.

Levi, Susannah V., Stephen J. Winters, and David B. Pisoni. 2007. Speaker-independent factors affecting the perception of foreign accent in a second language. *Journal of the Acoustical Society of America* 121: 2327–2338.

Lléo, Conxita, and Irene Vogel. 2004. Learning new segments and reducing domains in German L2 phonology. The role of the Prosodic Hierarchy. *International Journal of Bilingualism* 8: 79–104.

Long, Michael H. 1990. Maturational constraints on language development. *Studies in Second Language Acquisition* 12: 251–285.

Missaglia, Federica. 1999. Contrastive prosody in SLA – An empirical study with adult Italian learners of German. In *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 1, 551–554, 1–7 Aug 1999.

Missaglia, Federica. 2007. Prosodic training for adult Italian learners of German: The Contrastive Prosody Method. In *Prosodic description and teaching practice*, ed. Jürgen Trouvain and Ulrike Gut, 237–258. Berlin: Mouton de Gruyter.

Moyer, Alene. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21: 81–108.

Moyer, Alene. 2004. *Age, accent and experience in second language acquisition*. Clevedon: Multilingual Matters.

Moyer, Alene. Forthcoming. *Foreign accent: The phonology of non-native speech*. Cambridge: Cambridge University Press.

Munro, Murray J. 1998. The effects of noise on the intelligibility of foreign-accented speech. *Studies in Second Language Acquisition* 20: 139–154.

Munro, Murray J., and Tracey M. Derwing. 1995a. Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97.

Munro, Murray J., and Tracey M. Derwing. 1995b. Processing time, accent and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38: 289–306.

Munro, Murray J., and Tracey M. Derwing. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48: 159–182.

Nelson, Gayle L. 1991. Effective teaching behavior for international teaching assistants. In *Preparing the professoriate of tomorrow to teach: Selected readings in TA training*, ed. Jody D. Nyquist, Robert D. Abbott, Donald H. Wulff, and Jo Sprague, 427–434. Dubuque: Kendall Hunt.

Neufeld, Gerald G. 1978. On the acquisition of prosodic and articulatory features in adult language learning. *The Canadian Modern Language Review* 34: 163–174.

Oyama, Susan. 1976. A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research* 5: 261–283.

Piske, Thorsten. 2008. Phonetic awareness, phonetic sensitivity and the second language learner. In *Encyclopedia of language and education*, 2nd ed. Vol. 6, *Knowledge about language*, ed. Jasone Cenoz and Nancy H. Hornberger, 155–166. New York: Springer.

Piske, Thorsten, James E. Flege, Ian R.A. Mackay, and Diane Meador. 1999. Non-natives' production of vowels in conversational speech. *Collected papers from the 137th meeting of the Acoustical Society of America and the 2nd convention of the European Acoustics Association: Forum Acusticum*. Paper no. 2ASCB_4.

Piske, Thorsten, Ian R.A. MacKay, and James E. Flege. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29: 191–215.

Piske, Thorsten, James E. Flege, Ian R.A. Mackay, and Diane Meador. 2011. Investigating native and non-native vowels produced in conversational speech. In *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, ed. Magdalena Wrembel, Malgorzata Kul, and Katarzyna Dziubalska-Kołaczyk, 195–205. Frankfurt am Main: Peter Lang.

Purcell, Edward T., and Richard W. Suter. 1980. Predictors of pronunciation accuracy: A re-examination. *Language Learning* 30: 271–287.

Schumann, John. 1975. Affective factors and the problem of age in second language acquisition. *Language Learning* 25: 209–235.

Scovel, Thomas. 1969. Foreign accents, language acquisition, and cerebral dominance. *Language Learning* 19: 245–253.

Settinieri, Julia. 2008. Teaching pronunciation. In *Bielefeld introduction to applied linguistics: A course book*, ed. Stephan Gramley and Vivian Gramley, 1–12. Bielefeld: Aisthesis Verlag.

Steinlen, Anja K. 2005. *The influence of consonants on native and non-native vowel production: A cross-linguistic study*. Tübingen: Gunter Narr.

Thompson, Irene. 1991. Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning* 41: 177–204.

Trofimovich, Pavel, Patsy M. Lightbown, Randall H. Halter, and Hyojin Song. 2008. Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition* 31: 609–639.

Wennerstrom, Ann. 1998. Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition* 20: 1–25.

Winitz, Harris, Brenda Gillespie, and Jennifer Starcev. 1995. The development of English speech patterns in a 7-year-old Polish-speaking child. *Journal of Psycholinguistic Research* 24: 117–143.

Yule, George, and Doris Macdonald. 1994. The effects of pronunciation training. In *Pronunciation pedagogy and theory: New views, new directions*, ed. Joan Morley, 111–118. Alexandria: TESOL.

Chapter 5
Function *vs*. Form in Speech Prosody – Lessons from Experimental Research and Potential Implications for Teaching

Yi Xu

5.1 Introduction

Prosody is an indispensable part of a language and so its mastery is highly desirable for learners of any language, especially those at a relatively advanced level. But the pressing question is how this can be achieved, as prosody is notoriously difficult to learn (Atoye 2005; Dankovicova et al. 2007), and so far there has been a lack of effective ways to teach it (Atoye 2005). This difficulty is closely linked to the fundamental question of what is it that needs to be learned in terms of the prosody of a language. To start with an example, Fig. 5.1 displays the pitch tracts (dotted lines) of two sentences in General American English spoken by a female native speaker. The top one is a statement and the bottom one a question. In both sentences, the speaker puts emphasis on "Bloomingdales", which means that the word is the location of the nucleus (British) or nuclear accent (Pierrehumbert 1980). Whatever follows the large pitch movement should then be considered as the tail. As for what kind of nuclear it is, the top one could be considered as rise-fall and the bottom one low rise. The question is, however, are these the patterns that learners of English need to be taught?

To answer this kind of question, experimental findings on tone and intonation will first be reviewed, and then potential implications of these findings for the teaching of English as a second language will be explored.

Y. Xu (✉)
Division of Psychology and Language Sciences, Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK
e-mail: yi.xu@ucl.ac.uk

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_5,

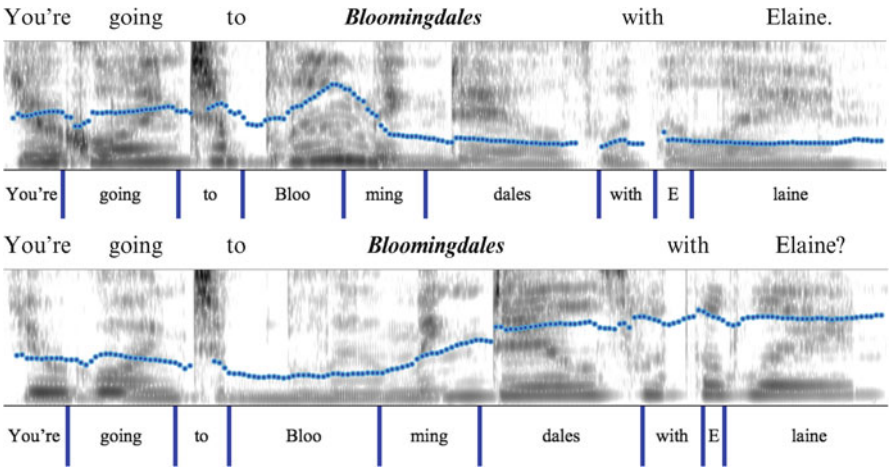


Fig. 5.1 Spectrograms and F_0 tracks of the sentence "You're going to Bloomingdales with Elaine", spoken by a female American English speaker as either a statement (**a**) or a question (**b**) (Samples from Liu and Xu 2007)

5.2 Basic Articulatory Mechanism – Lessons from Tone Languages

Figure 5.2a displays mean F_0 contours of four Mandarin lexical tones said in isolation, averaged across 40 repetitions by eight male speakers. These contours can be considered as close to the canonical forms of these tones, i.e., the underlying patterns, because they are free of influence from surrounding tones. In Fig. 5.3, the same four tones are produced in a five-syllable sentence by four male speakers. In the second syllable of each graph, we can see that the contours of the four tones bear much resemblance to those in Fig. 5.2a, with the exception of the L tone whose final rise in Fig. 5.2a is missing in Fig. 5.3, presumably due to a phonological rule, as will be discussed later (Chao 1968). Syllable 3 in Fig. 5.3, however, shows rather different F_0 contours despite the fact that its tone remains constant: H, R and F in (a), (b) and (c), respectively. These variations appear to be directly related to the fact that the F_0 contours of the four preceding tones have very different endings, and the beginning portion of the tone in syllable 3 seems to be a direct continuation of those endings. Interestingly, however, by the end of syllable 3, the four contours have virtually converged to a straight line whose height and slope are consistent with canonical forms of the tones: high-level for H, rising for R and falling for F. Similar patterns of contextual F_0 variations have been found in other tone languages (Gandour et al. (1994) for Thai, Wong (2006) for Cantonese and Laniran and Clements (2003) for Yoruba). From these studies we can see the following mechanisms of tone production:

1. A tone is produced by approaching its underlying canonical target *within* the syllable that carries it, i.e., from the syllable onset to the offset. This means that

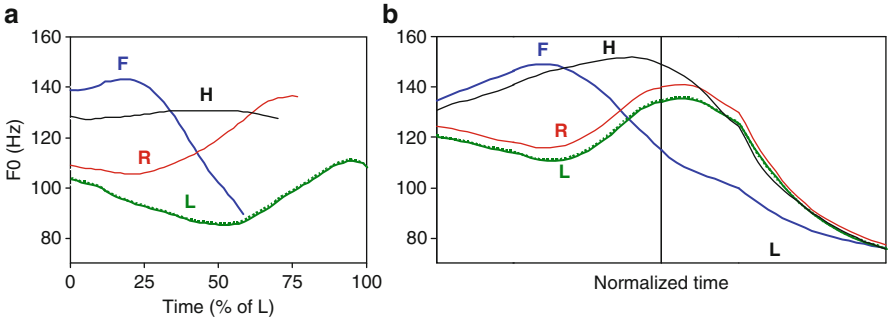


Fig. 5.2 (**a**): Four Mandarin tones produced in isolation, (**b**): Mandarin L tone after four different tones, produced in carrier phrases (Adapted from Xu 1997)

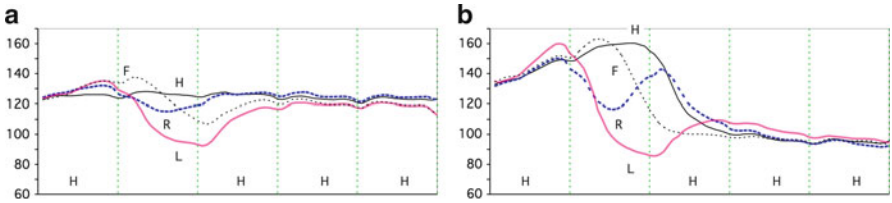


Fig. 5.3 Mandarin tone F following four different tones. (**a**): no narrow focus in the sentence; (**b**): focus on the F-carrying syllable. Each curve is an average of 20 tokens produced by four male speakers (five repetitions per speaker) (Data from Xu 1999)

the tone is articulated in synchrony with its host syllable. Such synchrony is maintained regardless of whether the initial consonant of the syllable is voiced or whether the syllable has a coda consonant (Wong and Xu (2007) for Cantonese; Xu and Xu (2003) and Xu (1998) for Mandarin).

2. It takes a substantial amount of time for F_0 to go from the ending value of the preceding tone to the desired value of the current tone. This is because, according to Sundberg (1979) and Xu and Sun (2002), at least 100 ms is needed to make a pitch change of even the smallest size, and the amount of time further increases with the size of the pitch change. Based on data from Xu (1997, 1999), the greater half of a syllable of an average duration of 200 ms is needed to make most of the tonal transitions. Furthermore, for the F tone in syllable 3 in Fig. 5.3a, because two pitch movements are involved within a single syllable, there are varying degrees of undershoot of the tonal target depending on the tone of syllable 2.
3. As a consequence of these basic tone production mechanisms, F_0 movements in the early half of the syllable mostly serve as transitions to the underlying tonal targets. And as such, they cannot be taken as the underlying tonal contours *per se*. The F_0 contours toward the end of the syllable, in contrast, seem to best resemble the underlying tonal targets.

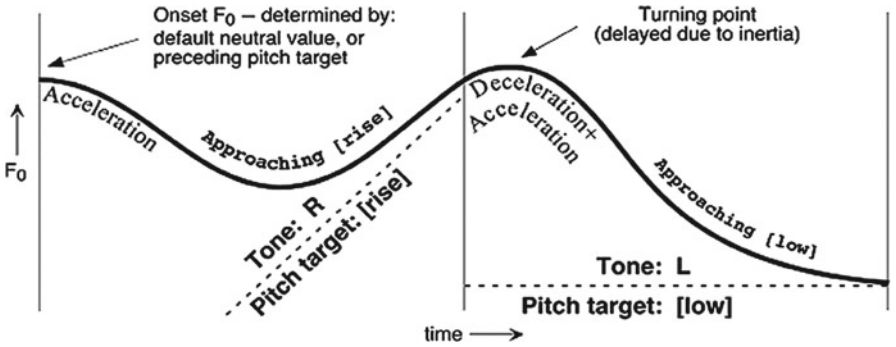


Fig. 5.4 Target approximation model (Adapted from Xu and Wang 2001)

The basic mechanism of tone articulation has been summarized in the Target Approximation (TA) model, as illustrated in Fig. 5.4 (Xu and Wang 2001). According to the model, lexical tones are produced by a process of articulatorily approaching successive local pitch targets, each in synchrony with the host syllable. In this model, there are several parameters that can be controlled by the speaker: target height, target slope and target strength (which defines the speed at which the target is approached), pitch range (which defines the overall height and vertical span of the target). The approximation of the target is always synchronized with the entire syllable, i.e., commencing with the syllable onset and terminating with the syllable offset (regardless of whether the target has been reached). Thus, the synchronization of pitch target and the syllable is assumed to be obligatory, leaving no room for the speaker to adjust the timing of the target relative to the temporal interval of the syllable. Finally, the TA model assumes that the final F_0 state of a syllable is transferred across the syllable boundary to become the initial state of the next syllable, as illustrated in Fig. 5.4.

5.3 Relevance for Non-tone Languages

One could argue that the articulatory mechanisms of tone production is applicable only to tone languages, and they are necessitated by the fact that in a language like Mandarin or Cantonese, every syllable is specified with a lexical tone and it is therefore articulatorily more demanding than a non-tone language like English (Ladd 1996). There is evidence, however, that these mechanisms also apply to non-tone languages. Xu and Sun (2002) examined native speakers of both Mandarin and English and found virtually no difference between the speakers of the two languages in terms of maximum speed of pitch change. Thus, native Mandarin speakers do not change pitch faster than native English speakers despite years of speaking a tone language. In the same study, previously reported data from Dutch and English were also reviewed and it was concluded that speakers of those two languages often

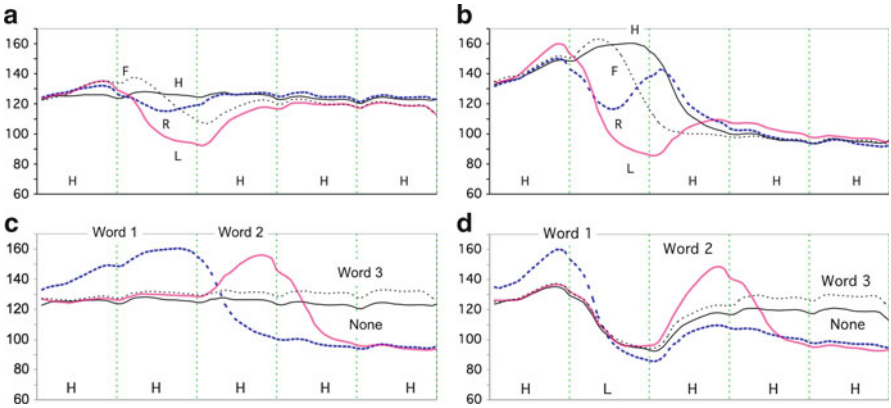


Fig. 5.5 (**a–b**): Time-normalized mean F_0 contours (20 tokens by four male speakers) of four five-syllable sentences in Mandarin, which differ in terms of the tone of the second syllable. In (**a**) there is no narrow focus in the sentence; in (**b**) the first two syllables are in focus. (**c–d**): two sentences with tone sequences of HHHH and HLHHH with focus on word 1 (syllables 1–2), word 2 (syllable 3) or word 3 (syllables 4-5) (Data from Xu 1999)

approach the maximum speed of pitch change in their speech (Caspers and van Heuven 1993). There is also evidence, either direct or indirect, that pitch production is synchronized with the syllable in English and other non-tone languages. Xu and Wallace (2004) show that F_0 transitions similar to those in Fig. 5.3 start from the syllable onset in American English regardless of whether the initial consonant is voiced or voiceless. Similar evidence was reported earlier by Silverman (1986). Further evidence of syllable-synchronization of pitch production as well as perception has been shown by Gao and Xu (2010) for Southern British English, Dilley and Brown (2007) for American English and Niebuhr (2007) for German. The existing evidence therefore demonstrates that non-tone language speakers change pitch no more slowly than speakers of a tone language, and they also likely produce underlying pitch targets in synchrony with the syllable.

Assuming that the mechanisms of producing pitch events in English are similar to those of tone languages, interpretation of F_0 contours in English should also follow that of tonal contours. In Fig. 5.1a, for example, the F_0 rise in the first syllable of "Bloomingdales" appears very sharp. But because the rising movement starts from the syllable onset and ends before the syllable offset, it should not be considered as due to an underlying rising target. Rather, it should be more like the Mandarin H tone shown in Fig. 5.5a, which means that the underlying target of the stressed syllable in "Bloomingdales" is likely to be a static high, at least in this particular example. In contrast, in Fig. 5.5b we can see that the R tone in Mandarin, which presumably does have a genuine rising target, exhibits a clear F_0 rise only in the second half of the syllable, and this rise continues into the beginning part of the next syllable. This "peak delay" is likely due to articulatory inertia and is modelled by the TA model as resulting from transferring the final

state of the previous F_0 movement to the beginning of the next syllable, as shown in Fig. 5.4. Similarly, the F tone, which presumably has a falling target, exhibits a clear fall in the second half of the syllable, as shown in Fig. 5.5c. Finally, the F_0 in the second syllable of "Bloomingdales" shows a sharp fall. However, because the F_0 movement is virtually levelled off before the syllable offset, the underlying pitch target is unlikely to be a [fall], but rather more likely to be a relatively low static register, similar to the Mandarin H tone in syllable 4 in Fig. 5.5a, whose height is substantially lowered by focus, as will be discussed later. More importantly, the sharp fall right after the F_0 peak, because it mostly occurs in the early part of the unstressed syllable after "Bloo-", should not be considered as due to a falling target.

5.4 Function *Versus* Form

To speak is to convey meanings, and prosody is an important meaning carrier. In the traditional approaches, however, prosodic categories are primarily defined by their forms, while their functional connotations are usually left vague. In the nuclear tone tradition, the forms of various types of nucleus are clearly defined: Rise, Fall, Rise-fall, Fall-rise, High fall, Low fall, High rise, Low rise, etc. (Palmer 1922; Crystal 1969; Halliday 1967; O'Connor and Arnold 1961). Bolinger, who has put much emphasis on the importance of communicative values of prosody, also first defines the basic intonational components in terms of their forms, e.g., A accent, B accent, etc. (Bolinger 1986). In the Pierrehumbert model, from which the Tonal elements of the ToBI transcription system were derived, intonational components are deemed to be phonological (i.e., contrastive) without establishing their meanings. In her own words,

> In the literature, one can distinguish two approaches towards the problem of establishing which intonation patterns are linguistically distinct and which count as variants of the same pattern. One approach attacks the problem by attempting to deduce a system of phonological representation for intonation from observed features of F0 contours. After constructing such a system, the next step is to compare the usage of F0 patterns which are phonologically distinct. The contrasting approach is to begin by identifying intonation patterns which seem to convey the same or different nuances. The second step is to construct a phonology which gives the same underlying representation to contours with the same meaning, and different representations to contours with different meanings… *The work presented here takes the first approach, in fact, it stops at the first step in the first approach* [my emphasis]. (p. 59)

It is not the case that traditional approaches are unconcerned with functions. Rather, they typically treat the forms as the defining properties of intonational categories. Functions, in contrast, are viewed as simply accompanying the forms. Pierrehumbert and Hirschberg (1990), for example, made an attempt to identify the meanings associated with the form-based intonational units established in Pierrehumbert (1980). The meanings identified in the paper, however, are long-winded, vague and heavily overlapped with each other. In addition, probably more critically, tonal units such as

H and L are conceptualized as directly meaningful. For example, the following are descriptions of the meanings of the H* and L* pitch accents:

> The H* accents above and in utterances in general convey that the items made salient by the H* are to be treated as "new" in the discourse. (p. 289)
>
> The L* accent marks items that S intends to be salient but not to form part of what S is predicating in the utterance. (p. 291)

Note that the two aspects of the traditional approaches, namely, (a) giving form priority and (b) assigning meanings directly to intonational units, are rather different from how the meaning-form relation is viewed in the case of the segmental aspect of speech. A segmental contrast is defined, first and foremost, by whether it can distinguish words or grammatical functions. Those differences that do not distinguish one word from another, e.g., the many /r/ variants in English, are considered as allophones rather than separate phonemes. Therefore, lexical contrast, as a function, is the *defining property* rather than a *subordinate* or *accompanying property*. Secondly, lexical contrast itself does not carry any specific meaning, but only serves to distinguish lexical items from one another. The specific meanings of the words are defined in morphology rather than in phonology, and only occasionally, e.g., in the case of onomatopoeia, is there any direct link between phonetic form and lexical meaning. Similarly, in the lexical use of pitch by either lexical tone or lexical stress, pitch itself carries no specific meaning but only serves to distinguish words. Nevertheless, there is a further aspect that is more unique to tone and intonation but less of a concern with segments. That is, segmental contrasts serve to distinguish only lexical items. But tonal and intonational components carry multiple functions. The traditional approaches, however, do not usually try to keep the functions clearly separated from each other. The definitions for the H* and L* pitch accents cited above exemplify the problem. In both cases, salience, newness and predication should be separate functions. But they are put together as components of a single hypothetical prosodic unit, because the F_0 shape is treated as the defining property.

An important reason for the persisting practice of giving primacy to prosodic form is that identifying functional categories in prosody from acoustic signals is difficult. This is because there are at least three degrees of separation between surface prosodic forms and the communicative functions they encode (Xu 2004). The first degree of separation is due to the articulatory mechanisms discussed in the previous section. Because of these mechanisms, directly observed surface acoustic forms often do not resemble the relatively invariant underlying phonetic targets that are used to encode information. Only through systematically controlled experiments can the invariance of the underlying targets be observed. The second degree of separation is due to target reassignment, which is a process of changing the underlying targets depending on various factors such as phonetic context and prosodic functions. A case in point is the Mandarin L tone, which has a rising tail when produced in isolation or sometimes in a sentence-final position. But the tail is missing when the L is followed by any other tone, as can be seen in Fig. 5.3, and its absence cannot be explained by articulatory mechanisms. In addition, the L tone in Mandarin changes into a R tone when followed by another L tone. This change again has no

plausible articulatory explanation. Such target reassignment happens not only to lexical tones, but also to lexical stress as a function of other prosodic functions, as will be discussed subsequently. The third degree of separation is due to parallel encoding of multiple communicative functions, as will be elaborated next.

5.4.1 Lexical Versus Extra-lexical Functions

Because F_0 is the major acoustic correlate of lexical tones, it could be assumed that intonational use of pitch is very limited in a tone language. But this assumption can easily be dismissed by some basic facts about pitch perception and production. On the one hand, native listeners can identify the four Mandarin tones in their native language with pitch differences no larger than 0.5 semitones (or 4 Hz) (Klatt 1973). On the other hand, the natural pitch range of an average speaker is well over two octaves, i.e., 24 semitones (Honorof and Whalen 2005). Thus, there is plenty of room for the use of pitch to encode both tone and intonation in a tone language. This is exactly what is found in languages like Mandarin. Figure 5.5a displays time-normalized mean F_0 contours of four five-syllable sentences in Mandarin, which differ from each other only in terms of the tone of the second syllable. Clear differences can be seen due to the tone of the second syllable. In Fig. 5.5b, the same sentences are spoken with prosodic focus on the first word consisting of two syllables. The differences in F_0 due to the tones of the second syllable are clearly exaggerated: the high pitch becomes even higher and the low pitch even lower. The differences between the sentences in Figs. 5.5a and 5.6b therefore constitute coding for the prosodic focus. Thus, both tone and focus are effectively encoded in parallel in Mandarin.

In Fig. 5.5b we can also see that after the exaggerated F_0 contours of the first two syllables, the pitch level of all the subsequent syllables are lowered relative to that in Fig. 5.5a. This can be more clearly seen in Fig. 5.5c, d, in which focus varies from word 1, word 2, word 3 and none. It is obvious that as long as focus is not sentence-final, there is significant lowering and narrowing of post-focus pitch range. Chen et al. (2009) also find that post-focus intensity is significantly reduced. The reduction of pitch range and intensity after prosodic focus is known as *post-focus compression* (Chen et al. 2009) or PFC for short (Xu et al. in press).

Parallel encoding of lexical and extra-lexical functions has also been found in English, a non-tone language. Fry (1958) shows that listeners use the pitch difference between two adjacent syllables to determine which is lexically stressed. The syllable with the higher pitch is heard as stressed. He finds that a difference as small as 5 Hz leads to unambiguous judgment of lexical stress. What Fry has found is only about how lexical stress is used to distinguish words in English, because listeners in his experiments were asked to judge whether a word like “digest” or “permit” is a noun or a verb, but not whether any syllable is stressed. Thus, a 5 Hz difference in F_0 is sufficient to indicate the functional contrast of lexical distinction. Such a functional use of F_0 is actually very similar to lexical tone, and the similarity is especially high in comparison with the neutral tone in Mandarin (Chen and Xu 2006).

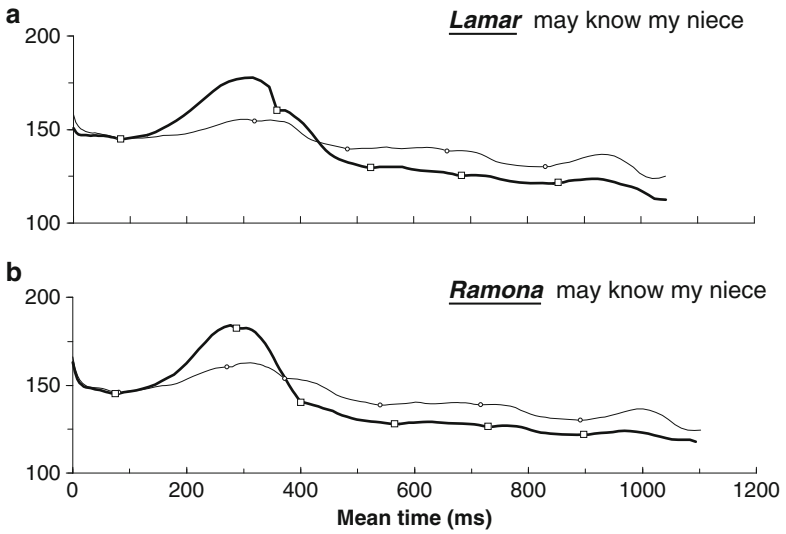


Fig. 5.6 Time-normalized mean F_0 contours (each over 49 tokens by seven speakers) of two sentences in American English. *The thin curves* are from sentences said without prosodic focus; *the thick curves* are from sentences with focus on the first word (Adapted from Xu and Xu 2005)

Also similar to Mandarin, the focus function is overlaid on the lexical function (Cooper et al. 1985; Pell 2001; Xu and Xu 2005). Figure 5.6 shows mean F_0 contours of two English sentences spoken by seven native American English speakers. The thin curves are from sentences said without prosodic focus, in which the small F_0 undulations correspond to the relative stress of the individual syllables. When a word is focused, as shown by the thick curves, the F_0 contour of the focused word is increased, but the increase does not seem to create any new peaks in addition to those already in the neutral-focus curve. Thus, focus seems to expand the pitch range of the focused words relative to that of the neutral focus F_0. Furthermore, the F_0 of the post-focus words is lowered, but again, the lowering does not seem to eliminate F_0 peaks from the neutral-focus F_0 curves. Xu and Xu (2005) found that the height of these small F_0 "bumps" is comparable to the minimal F_0 difference needed for the perception of lexical stress according to Fry's (1958) study. So the pitch range of post-focus words is reduced, just like in Mandarin. Therefore, we can see that lexical and focal contrasts are both encoded with F_0, and the encoding of the two is done in parallel, allowing both to be sufficiently distinctive.

More interestingly, again like Mandarin and many other Chinese languages, target reassignment also occurs in English, although the conditions for its occurrence are different. One of the conditions that triggers target reassignment is focus. In Fig. 5.6a the final portion of the F_0 contour levels off toward the end of the stressed syllable /mar/ in "Lamar" when the sentence has no narrow focus (thin line). When "Lamar" is focused, however, the final portion of /mar/ becomes falling. The fall is not as sharp as that of the F tone in Mandarin shown in Fig. 5.5 but sharper than that

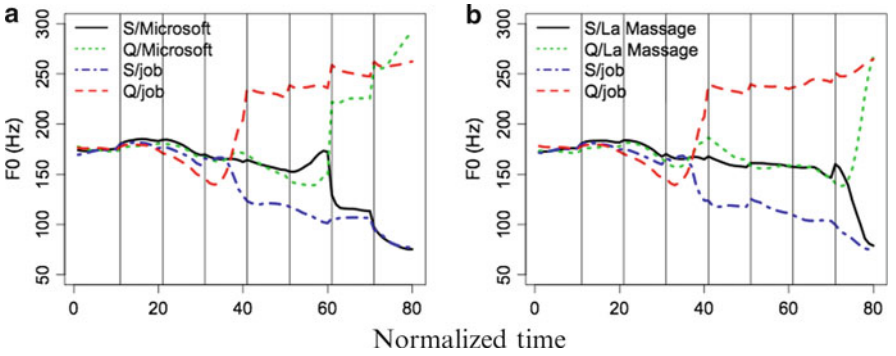


Fig. 5.7 Mean F_0 contours of American English statements and questions with focus. The word after "/" is focused. *S* statement, *Q* question. *The vertical lines* indicate syllable boundaries. (**a**) You want a job with Microsoft, (**b**) You want a job with La Massage (Data from Liu and Xu 2007)

of the Mandarin High tone shown in Fig. 5.5. More detailed examination shows that individual English speakers differ in terms of how sharp the fall is. But overall, there seems to be a tendency to change the targeted F_0 trajectory of a word-final stressed syllable from [high] to [fall]. This focus-triggered target change is highly condition-specific, because it does not happen to stressed syllables that are not word final. In Fig. 5.6b, for example, F_0 continues to rise toward the end of the stressed syllable /mo/ in "Ramona", whether or not the word is in focus.

Thus, in both Mandarin and English, F_0 is used to simultaneously encode lexical and focal contrasts. The two are encoded in parallel by employing different encoding strategies. Lexical contrast is encoded mainly with syllable-sized pitch targets, while focal contrast is encoded mainly by modifying the pitch range of the local targets. For English, however, focus also interacts with lexical stress by changing the local targets of word-final stressed syllables.

5.4.2 *Encoding of Sentence Type Together with Lexical Stress and Focus*

Beside lexical stress and focus, another function that further shapes English prosody is sentence type, which determines whether an utterance is spoken as a statement or a question. This is achieved again through parallel encoding and target reassignment (Eady and Cooper 1986; Liu and Xu 2007), as illustrated in Fig. 5.7. The word "job" in Fig. 5.7a, b and the stressed final syllable of "massage" in Fig. 5.7b both show a sharp F_0 fall toward the end of the syllable when they are under focus in a statement (dashed line), which is consistent with what was described in the previous section.

But when the sentences are questions, these syllables all show a sharp F_0 rise all the way to the end of the syllable. This indicates that the underlying pitch target of the syllable, which is stressed and word-final, changes from falling to rising. In addition, Liu and Xu (2007) find that the pitch targets of non-word-final stressed syllables also tend to have a rising target, although the slope of the rise is not nearly as steep as that of word-final stressed syllables. This can be seen in the first syllable of "Microsoft" in Fig. 5.7a.

In addition to the local F_0 excursions, Fig. 5.7 also shows that the height of the F_0 contour is dramatically increased immediately after the focused stressed syllable in a question, and it continues to rise all the way to the end of the sentence. This pattern contrasts with the drastic post-focus drop of F_0 height in a statement. Meanwhile, the post-focus local pitch movements are very small in both statements and questions. It appears that post-focus compression of pitch range in terms of local excursions applies in English whether the sentence is a statement or question, but the overall height is very different, high in a question but low in a statement. In other words, the overall height of post-focus pitch range is functionally determined by sentence type, whereas the magnitude of the local excursions is closely related to focus.

To summarize, much of the variations in English result from the parallel encoding of three basic communicative functions: lexical stress, prosodic focus and sentence type, with the following rules:

1. Each syllable, whether stressed or unstressed, is assigned a local pitch target; the properties of the target are jointly determined by lexical stress, focus and sentence type;
2. In a statement a stressed syllable is assigned a high target unless it is word final and on-focus, in which case it is assigned a falling target;
3. In a question, a stressed syllable is assigned a rising target;
4. Prosodic focus expands the pitch range of the focused word and compresses the pitch range of all post-focus words, but leaves the pitch range of pre-focus words unchanged from that of the neutral focus sentence;
5. A statement gradually lowers the pitch range throughout the sentence, but adds an extra drop immediately after the stressed syllable on focus and at the end of the sentence;
6. A question does the opposite to the pitch range, i.e., raising it throughout the sentence, but adding an extra upward boost immediately after the focused stressed syllable as well as at the end of the sentence.
7. Unstressed syllables are not targetless but are assigned a mid target with weak articulatory strength (Xu and Xu 2005).

5.4.3 Additional Functions

Beside lexical stress, focus and sentence type, there are other communicative functions that are also encoded mainly through prosody. So far, however, there has not

yet been empirical research that has generated highly specific descriptions of their patterns. One of the functions is topic or turn initiation, which raises the pitch range of the beginning of a sentence (Lehiste 1975). Another is known as the contradiction contour, which exhibits a global fall-rise pattern across the whole sentence (Liberman and Sag 1974). The exact condition of its occurrence is not yet fully clear, however, making it difficult to be experimentally investigated. There are other possible stylistic global patterns described in the descriptive literature (Bolinger 1986, 1989; Cruttenden 1997; Crystal 1969; Halliday 1967; O'Connor and Arnold 1961). But again, the exact function and condition of occurrence of those patterns need to be systematically investigated. Much more empirical research is therefore needed.

5.5 Implications for Teaching English Intonation

The empirical findings on tone and intonation discussed so far may suggest alternative approaches to the teaching of English intonation that are different from the current practices. The dominant strategy in teaching English intonation, to my knowledge, is based the nuclear tone tradition. As discussed earlier, this tradition is mostly form-oriented. The fact that its effectiveness has yet to be shown (Atoye 2005; Currie 1980) demonstrates that new methodology could be considered. Since I am unaware of any empirical research on teaching English intonation based on the new findings, I can only offer some preliminary suggestions.

The first suggestion is that it could be beneficial to teach learners of English *functionally defined* intonation patterns. For example, the summary of how lexical, focal and sentential functions are conveyed through pitch, as discussed in previous sections, could be developed into possible teaching instructions:

1. In a statement, a stressed syllable should have higher pitch than an unstressed syllable; its pitch contour is preferably level unless it is word-final and focused or sentence-final, in which case the contour is preferably falling.
2. In a question, a stressed syllable should have *lower* pitch than an unstressed syllable; its pitch contour is preferably slightly rising unless it is word-final and focused or sentence-final, in which case the contour should be sharply rising.
3. If a word is focused, the pitch of its stressed syllable should be exaggerated, i.e., becoming higher in a statement but lower in a question.
4. Immediately after the stressed syllable of a focused word, the pitch of all the following syllables in the sentence should be substantially lowered in a statement but raised in a question.

There are foreseeable difficulties in implementing these suggestions, however. The first is that learners may vary extensively in terms of their ability to follow instructions on both local pitch targets and global pitch patterns, as found by Dankovicova et al. (2007). As a result, many learners may not be able to follow instructions about pitch patterns. The second potential difficulty is the interference

of learners' first language. In particular, recent research has found that there is a typological divide among the world's languages in terms of the application of PFC, i.e., the extensive reduction of pitch range after focus, as described in Sect. 5.4.1 and mentioned again in suggestion 4 above. Languages in which PFC applies include Indo-European languages like English, German, Italian, Swedish, etc., Altaic languages like Turkish, Japanese and Korean, Uralic languages like Finnish, Semitic languages like Arabic and Northern Chinese languages like Mandarin (see Xu 2011 for a brief summary). Languages in which PFC does not apply include Southern Chinese languages like Cantonese and Taiwanese (Chen et al. 2009; Wu and Chung 2011), Mon-Khmer languages like Wa, Deang (Wang et al. 2011) and possibly Vietnamese (Jannedy 2007) and many African languages, including Sotho, Buli, Chichewa and Hausa, as summarized by Zerbian et al. (2010). Native speakers of these and probably many more "non-PFC" languages may thus have difficulty learning PFC, at least without explicit instructions, as indicated by the findings of a number of recent studies (Chen et al. 2009; Wu and Chung 2011).

Note, however, that these difficulties would occur even with traditional approaches to intonation teaching. A function-based approach, nevertheless, may potentially make it easier for learners to become aware of the most critical components of intonation that are functionally relevant. In any case, the strategy of teaching the interactive patterns of lexical stress, focus and sentence type has never been tested before, to my knowledge. Given the clear findings in the first language research, it is at least worth exploring.

A further suggestion that could be offered is in regard to the articulatory mechanisms of tone production discussed near the beginning of the chapter. That is, given that the obligatory synchronization of local pitch targets to the syllable, there is no need to teach learners the alignment of F_0 peaks and valleys. Such alignment has been much discussed in recent literature, based on the framework of the Autosegmental-Metrical Phonology of intonation (Pierrehumbert 1980), also known as Intonational Phonology (Ladd 1996). According to the Target Approximation model shown in Fig. 5.4, the exact location of F_0 turning points is a direct consequence of the underlying pitch target of the current syllable and those of the adjacent syllables. If the alignment of the underlying pitch targets is obligatory, there is little room for learning because learning is possible only when the learners have free choices. The real choices learners have are likely in terms of the properties of the pitch target in terms of height and slope, as shown in Fig. 5.4, but not in terms of its micro-timing within the syllable.

5.6 Concluding Remarks

This chapter has provided a brief overview of the latest findings on tone and intonation from an articulatory-functional perspective. It is shown in particular that much of the English intonation can be understood in terms of parallel encoding of multiple communicative functions, including lexical stress, focus and sentence type.

Based on these findings, some suggestions are offered in terms of possible alternative teaching strategies that depart from current common practice. In general terms, first, it might be more effective to teach functionally defined prosodic patterns rather than patterns classified in terms of their surface form. Second, it might be more effective to teach syllable-based pitch targets rather than word- or phrase-based whole-contours or their alignments. Finally, it might be more effective to teach complex prosodic forms as resulting from interactions of multiple communicative functions, each with a relatively simple underlying form rather than as prosodic gestalts, each with a convoluted set of meaning attributes. Whether these suggestions will lead to improvement in effective teaching of English intonation, however, awaits empirical research that puts them to real test.

References

Atoye, R.O. 2005. Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies* 14: 26–42.

Bolinger, D. 1986. *Intonation and its parts: Melody in spoken English*. Palo Alto: Stanford University Press.

Bolinger, D. 1989. *Intonation and its uses – Melody in grammar and discourse*. Stanford: Stanford University Press.

Caspers, J., and V.J. van Heuven. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50: 161–171.

Chao, Y.R. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Y., and Y. Xu. 2006. Production of weak elements in speech – Evidence from f_0 patterns of neutral tone in standard Chinese. *Phonetica* 63: 47–75.

Chen, S.-w., B. Wang, and Y. Xu. 2009. Closely related languages, different ways of realizing focus. In *Proceedings of Interspeech 2009*, Brighton, UK, 1007–1010.

Cooper, W.E., S.J. Eady, and P.R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America* 77: 2142–2156.

Cruttenden, A. 1997. *Intonation*. Cambridge: Cambridge University Press.

Crystal, D. 1969. *Prosodic systems and intonation in English*. London: Cambridge University Press.

Currie, K.L. 1980. An initial search for tonics. *Language and Speech* 23: 329–350.

Dankovicova, J., J. House, A. Crooks, and K. Jones. 2007. The relationship between musical skills, music training and intonation analysis skills. *Language and Speech* 50: 177–225.

Dilley, L.C., and M. Brown. 2007. Effects of pitch range variation on f0 extrema in an imitation task. *Journal of Phonetics* 35(4): 523–551.

Eady, S.J., and W.E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402–416.

Fry, D.B. 1958. Experiments in the perception of stress. *Language and Speech* 1: 126–152.

Gandour, J., S. Potisuk, and S. Dechongkit. 1994. Tonal coarticulation in Thai. *Journal of Phonetics* 22: 477–492.

Gao, H., and Y. Xu. 2010. Ambisyllabicity in English: How real is it? In *Proceedings of The 9th Phonetics Conference of China (PCC2010)*, Tianjin.

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

Honorof, D.N., and D.H. Whalen. 2005. Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America* 117: 2193–2200.

Jannedy, S. 2007. Prosodic focus in Vietnamese. In *Interdisciplinary Studies on Information Structure (ISIS),* ed. S. Ishihara, S. Jannedy and A. Schwarz, 209–230. Potsdam: Potsdam University Press.

Klatt, D.H. 1973. Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *Journal of the Acoustical Society of America* 53: 8–16.

Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Laniran, Y.O., and G.N. Clements. 2003. Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics* 31: 203–250.

Lehiste, I. 1975. The phonetic structure of paragraphs. In *Structure and process in speech perception*, ed. A. Cohen and S.E.G. Nooteboom, 195–206. New York: Springer.

Liberman, M., and I. Sag. 1974. Prosodic form and discourse function. In *Proceedings of Papers from the 10th Regional Meeting*, 416–427. Chicago: Chicago Linguistic Society.

Liu, F., and Y. Xu. 2007. Question intonation as affected by word stress and focus in English. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1189–1192.

Niebuhr, O. 2007. The signalling of German rising-falling intonation categories – The interplay of synchronization, shape and height. *Phonetica* 64: 174–193.

O'Connor, J.D., and G.F. Arnold. 1961. *Intonation of colloquial English*, 2nd ed. London: Longmans.

Palmer, H.E. 1922. *English intonation, with systematic exercises*. Cambridge: Heffer.

Pell, M.D. 2001. Influence of emotion and focus on prosody in matched statements and questions. *Journal of the Acoustical Society of America* 109: 1668–1680.

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. PhD dissertation, MIT, Cambridge, MA. [Published in 1987 by Indiana University Linguistics Club, Bloomington].

Pierrehumbert, J., and J. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, ed. P.R. Cohen, J. Morgan, and M.E. Pollack, 271–311. Cambridge: MIT Press.

Silverman, K. 1986. F_0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica* 43: 76–91.

Sundberg, J. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7: 71–79.

Wang, B., L. Wang, and T. Kadir. 2011. Prosodic encoding of focus in six languages in China. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 144–147.

Wong, Y.W. 2006. Contextual tonal variations and pitch targets in cantonese. In *Proceedings of speech prosody 2006*, Dresden, Germany, PS3-13-199.

Wong, Y.W., and Y. Xu. 2007. Consonantal perturbation of f_0 contours of Cantonese tones. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrucken, 1293–1296.

Wu, W.L., and L. Chung. 2011. Post-focus compression in English-Cantonese bilingual speakers. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 148–151.

Xu, Y. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25: 61–83.

Xu, Y. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55: 179–203.

Xu, Y. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27: 55–105.

Xu, Y. 2004. The PENTA model of speech melody: Transmitting multiple communicative functions in parallel. In *Proceedings of from sound to sense: 50+ years of discoveries in speech communication*, Cambridge, MA, C-91-96.

Xu, Y. 2011. Post-focus compression: Cross-linguistic distribution and historical origin. In *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 152–155.

Xu, Y., and X. Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399–1413.

Xu, Y., and A. Wallace. 2004. Multiple effects of consonant manner of articulation and intonation type on F_0 in English. *Journal of the Acoustical Society of America* 115(Pt. 2): 2397.

Xu, Y., and Q.E. Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33: 319–337.

Xu, C.X., and Y. Xu. 2003. Effects of consonant aspiration on Mandarin tones. *Journal of the International Phonetic Association* 33: 165–181.

Xu, Y., and C.X. Xu. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics* 33: 159–197.

Xu, Y., S.-w. Chen, and B. Wang. in press. Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *To appear in The Linguistic Review*.

Zerbian, S., S. Genzel, and F. Kügler. 2010. Experimental work on prosodically-marked information structure in selected African languages (Afroasiatic and Niger-Congo). In *Proceedings of Speech Prosody 2010*, Chicago, 100976: 1–4.

Chapter 6
Prosodic Adaptation in Language Learning

Marie Nilsenová and Marc Swerts

6.1 Introduction

If one compares a spoken interview between different dialogue partners with a transcript of the same interaction, quite often the impression arises that the written version does not do full justice to the richness of the spoken interaction. Indeed, when listening to a recording of a spoken conversation, it becomes immediately clear that the interlocutors do more than simply exchange sequences of sentences. Spoken messages as produced in real interactions usually have a certain "*je-ne-sais-quoi*" which allows listeners to infer easily their semantic and pragmatic meaning, compared to how they would process written records of the same dialogue exchange. That extra information is multifarious. For instance, the way people speak can reveal who the dominant person in a specific conversational setting is or whether someone is ironic about the things he or she claims; alternately, it becomes clear whether the produced utterances were intended as simple statements or rather as questions and it is often striking to observe how fluently dialogue partners take turns without too much overlap and with minimal delay between speaking turns. These are just a few examples that show that one really has to be able to "hear" a speaker to fully appreciate all the connotations of his or her messages and their function in the conversation.

There is a whole body of literature suggesting that this "extra" information, which is not necessarily contained in the words or syntactic structures of sentences, is largely signaled by prosody. Roughly speaking, one could define prosody as the whole gamut of features that do not so much determine what speakers say but rather, how they say it. It includes features such as intonation (speech melody), tempo, rhythm, loudness and specific variations in voice quality. All natural languages spoken

M. Nilsenová (✉) • M. Swerts
Department of Communication and Information Sciences, Tilburg centre for Cognition and Communication, Tilburg University, Tilburg, The Netherlands
e-mail: m.nilsenova@uvt.nl

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_6,

in the world have prosody; there are no languages in which utterances are produced at a constant tempo, in a straight monotone and with no variation in loudness or voice quality. If we tried to produce utterances without any prosodic variation, we would probably resemble robots in a science-fiction movie.

Given that prosody is a natural ingredient of all languages, speakers obviously need to learn to produce utterances with appropriate prosodic patterns, both in their first language as well as in the foreign languages they acquire. From the perspective of (second) language learning, a few questions then arise, such as: does a growing child need to learn prosodic features of a language or is a child genetically or physiologically predisposed to produce such features? To what extent are languages similar or different with respect to the implementation of prosodic properties? And how can possible prosodic differences between languages best be described? In the following sections, we aim to shed light on these issues. We first discuss some arguments about the extent to which the acquisition of prosody is a matter of nature or nurture. Subsequently, we focus on the role of prosodic adaptation in language acquisition and, especially, the role it plays in creating social rapport. We end with some recommendations for doing prosody research, especially in view of didactic programs in second-language acquisition.

6.2 Nature or Nurture?

6.2.1 A Question of Nature

Prosody is among the first language phenomena experienced by a child. Even before they are born, children are exposed to the melodic patterns and rhythms of their mother's language. Though at this stage semantic interpretation is mostly lacking, it has been shown that babies just a few weeks old show a preference for the language with which they were surrounded prior to birth. In the subsequent pre-linguistic period, prosody continues to play an important role in helping to distinguish between speech and non-speech sounds (such as laughing) and in supporting speech segmentation (Brent and Siskind 2001; Soderstrom et al. 2008). The importance of prosody in the early stages of language acquisition thus suggests that the recognition of prosodic features is inborn and universal, in that infants are able to make use of prosodic information early on.

In fact, we know that prosody has a clear biological basis. Unlike pronunciation or phoneme inventory, the prosodic properties of a speaker's voice are largely determined by body size, age and emotional state. For example, shorter speakers tend to have shorter vocal tracts and thus produce higher frequencies of speech formants and faster decays (Irino and Patterson 2002). Longer and heavier vocal cords give rise to a lower glottal pulse rate and thus, lower pitch (compare the pitch of male and female speakers or the pitch of mature speakers and children; von Kriegstein et al. 2006; Kunzel 1989). From a behavioral point of view, respiration and muscle tension

are directly influenced by the speaker's emotional state and determine the nature of the vocal expression (Scherer 1989; Juslin and Laukka 2003). In sum, there are a number of physiological characteristics that determine the prosodic properties in speakers' speech production; these, in turn, give rise to prosodic universals.

Similarly to word order effects, it has been shown that languages share some universal tendencies in the use of pitch and intensity. A famous example is the widespread tendency for questions to be signaled with rising intonation (70% of languages in a sample of 250; Bolinger 1978). This universal is generally referred to by the term 'Frequency Code' (Ohala 1983, 1984) and it was extensively examined in the work of Gussenhoven (2002, 2004) and his collaborators. Apart from the Frequency Code, Gussenhoven identified two other 'biological' codes related to the use of intonation for signaling linguistic and paralinguistic information. In sum, the three codes, to wit the Frequency Code, the Effort Code and the Production Code, are based on the following observations (Gussenhoven 2002):

1. *The Frequency Code*: The larynxes of smaller and/or younger speakers contain lighter and smaller vocal cords, which vibrate faster (with higher frequency and thus higher pitch), compared to larger/mature speakers. As a consequence, lower pitch became associated with dominance. The difference between high/low pitch can be exploited to express (un)certainty, (un)friendliness and dominance/submissiveness.
2. *The Production Code*: The breathing process determines the amount of energy in speech. Naturally, high pitch and high intensity are associated with the beginnings of utterances, while low pitch and intensity are associated with the ends. Speakers make use of this fact to signal turn yielding or turn keeping.
3. *The Effort Code*: The amount of energy expended on speech production varies – putting in more effort will lead to more precise articulatory movements and to more canonical and more numerous pitch movements. Bestowing energy on the production process results in less slurring and undershooting of targets.

The Frequency Code is based on Ohala's claim that there is a correlation between larynx size, fundamental frequency and body size, which is used for expressing power relations. The code's paralinguistic connotations are that of being submissive, feminine, polite, vulnerable and friendly (for high pitch), as opposed to dominant, masculine, confident and aggressive (for low pitch). On the level of linguistic information, higher pitch is supposed to convey uncertainty and hence, questioning, based on the idea that "when asking questions, one is dependent on the other's good will for the information requested" (Chen 2004:33). The Production Code is linguistically exploited to signal new topics (high pitch), as opposed to topic continuations (low pitch). The basis for the code is sometimes referred to as 'final lowering' (Liberman and Pierrehumbert 1984). At least in some cases, the Production Code can have an opposite tendency to the Frequency Code because it predicts that speakers will have the predisposition to lower their pitch towards the end of utterances – a tendency that may conflict with their exploitation of the Frequency Code to signal information seeking. Finally, the Effort Code is claimed to be behind large pitch excursions associated with informationally salient items. Paralinguistically, it is

linked to the fact that active emotional states, such as joy or anger, result in dynamic speech production and increased loudness. These three prosodic universals or 'codes' are assumed to be phonetically implemented in the paralinguistic production and interpretation of pitch in all languages[1] and to have an effect on the phonology in most.[2]

Is it then at all necessary to learn L2 prosody? The above suggests that speakers and listeners should very naturally acquire diverse aspects of prosodic variation, especially if they can be related to biological or physiological factors. But while such determinants are likely to be universal and may therefore have similar effects on the prosodic structures of speakers across the globe, there also appear to be language-specific elements of prosodic structure that need to be "learned" by members of a specific language community. In particular, when discussing prosodic differences between languages, it makes sense to distinguish between differences in prosodic form and differences in the way these forms are linked to specific communicative or social functions. We will discuss both issues in the following section.

6.2.2 A Question of Nurture

When looking at the way speakers vary their pitch in the course of an utterance, it appears that they do not produce all the possible melodic patterns that their vocal apparatus is capable of: speakers can vary their pitch in a much more dramatic way (speed of change, jumps in pitch, pitch span) than they tend to do in a normal conversation. This suggests that such melodic patterns, as well as other prosodic variations, are constrained by a specific set of conventions. At first blush, speech variation at the prosodic level thus resembles that at the level of individual speech sounds, as it appears to be true for both levels of sound structure that the range of permissible variation in speech is limited by language-specific constraints. When talking, human beings make only limited use of their vocal capabilities, as they produce merely a subset of the sounds they are able to utter. The sound structure of a specific language is to a large extent dictated by a linguistic code, i.e., a specific ensemble of conventions a speaker shares with members of the same language community. For instance, at the segmental level, it appears that the sound/s/is part of the Dutch phonetic inventory, whereas/θ/is not; the latter is a phoneme in English, though. Similarly, prosodic variation is restricted within language-specific bounds.

[1]Chen (2004), however, examined the paralinguistic interpretation of the codes in a series of perceptual experiments and found varying tendencies in judgments of speakers of different languages.

[2]Overall, languages are expected to exhibit the tendency to use grammatical categories directly derived from the paralinguistic meaning. Even in those languages where the linguistic implementation of intonation categories goes against the Biological Codes (as seems to be the case, e.g., for rising statements in Northern Irish or falling questions in Hungarian), speakers will still exploit the original phonetic effect to convey their attitudes and emotions (Gussenhoven 2002).

6.2.2.1 Differences in Global Prosodic Conventions

There are reasons to assume that global prosodic conventions differ between languages. In the IPO tradition of intonation research, for instance, it has been argued that languages vary regarding the basic set of pitch rises and falls and in the way these can form larger melodic contours. According to this model, rises and falls in Dutch have very specific characteristics in terms of timing, slope and excursion size, with settings that are different from those in German, Russian or British English (Collier 1991). There are also specific constraints on how these language-specific pitch movements can be combined to form larger-scale melodic structures (Hart et al. 1990). In addition, languages can differ regarding overall settings in pitch range and register; it has, for instance, been argued that speakers of British English use a wider melodic span than speakers of Dutch (Willems et al. 1988).

Interestingly, the melodic dissimilarities may relate to other aspects of prosodic structure. One issue that has been a subject of debate is whether languages can differ in speech tempo. For example, it is generally assumed that speakers of Italian produce their speech at a comparatively fast tempo. However, research by Den Os (1988), who compared speech rate of Italian speakers with that of Dutch speakers, did not find very conclusive evidence that would support this assumption. Presumably, perceptual differences in speech tempo between these two languages are either a myth or are caused by the way prosodic information is processed by our perceptual system. In particular, our perception of "speed" might be affected by other aspects of prosodic structure, such as the melodic variation. As observed by Rietveld and Gussenhoven (1987), utterances that have the same objective rate (in terms of words per second) may sound faster or slower depending on whether the utterance has more or less melodic variation. Possibly, the perceptual differences in speech rate between Italian and Dutch might be related to the fact that these languages differ in intonational patterns, which, in turn, affect the perception of speech rate.

Differences in global prosodic conventions can be found not just between languages but even within a single language. For instance, work on Swedish shows that the varieties of Stockholm, Göteborg, Dalarna and Malmö can be quite distinct in the way speakers time pitch movements to distinguish specific accent types (Bruce and Thelander 2001). Along the same lines, pitch movements are differently aligned in a syllable for variants of German as, spoken in the North and South of Germany (Atterer and Ladd 2004) and for different dialects of Italian (Grice et al. 2005). At the temporal level, it has been shown that Dutch in the Netherlands is generally spoken at a "faster" rate compared to Dutch spoken in the Flemish part of Belgium, whereas within the Netherlands, speakers of the Limburgian part are slower than those living near Amsterdam (Verhoeven et al. 2004). Note that the regionally determined variation in prosodic structures may have consequences for learning programs that focus on aspects of prosodic structure. An important question that needs to be addressed for the purposes of L2 learning is which variety of a language, if any, serves as the norm within a specific community and what the role of prosodic variation is in that respect. The situation may be different for different language

communities because some languages may have clear linguistic standards, whereas this could be less true for others. For instance, l'Académie Française is generally regarded as the institution that prescribes what should be viewed as good or bad French; similarly, there is a consensus that the variant of English spoken by BBC newsreaders represents the best version of their language. However, such linguistic norms appear to be less dominant or less clearly defined in languages such as Swedish or Dutch. Accordingly, an L2 learner of Dutch who needs to master the prosodic structures of the language is faced with the problem of deciding which variant he or she should take as example.

6.2.2.2 Differences in Specific Prosodic Functions

In addition to the global differences discussed above, languages can also differ in the extent to which they use intonational and other prosodic features for linguistic and social purposes. One problem, however, is that our knowledge of functional aspects of prosodic variation is still limited. Studies so far clearly show that prosodic features are multi-functional, as they are exploited for a wide range of communicative functions. In this respect, they thus play a different role than the one served by segmental variation. Individual sounds, like/b/and/p/in Dutch or English, are employed as basic building blocks out of which meaningful units are constructed. Though they have no meaning on their own, they may change meaning in a discrete way, as the replacement of one phoneme by another can create a completely new word. For instance, when we change the initial/b/in "book" to a/t/, we get a new word ("took") with a completely different meaning. There are only a few examples of how variation at the prosodic level can change meaning in such a categorical manner. In the following famous example, due to Halliday (1967):

1. *John called Sam a republican, and then he insulted him,*

the meaning of the sentence changes fundamentally, depending on whether "insulted" gets a pitch accent, or rather the pronouns ("he", "him") before and after that verb. However, in many other cases, prosodic variation is usually not distinctive in this structural linguistic sense but, rather, has expressive power. Very often, prosodic variation does not change the truth-conditional properties of a sentence, but adds something "extra" to the content of the message that is not already conveyed by the lexical items and the syntax of the sentence. Nonetheless, the specific prosodic functions differ across languages, as can be illustrated on the phenomenon of prosodic highlighting.

Prosodic Highlighting

Not all information in a discourse is equally important. For instance, to study a written text, readers, when first exposed to a document that they have to learn, typically highlight some words that represent the main message by underlining them. If properly

done, at a later stage, readers do not need to read the whole text again but can focus on the words they highlighted to reconstruct the gist of the document. Where a writer and a reader can rely on typographic means (such as the layout of a text, use of boldface or italics) to identify the important parts of a text, a speaker-listener can exploit prosody for similar purposes, making use of the features already discussed with respect to the Effort Code. In particular, it has been shown that languages such as English or Dutch use pitch accents as markers of words that are informationally important. Pitch accents in English and Dutch can have quite different shapes, but usually consist in a relatively fast melodic rise, a fall, or a combination of these, which make a word sound as more prominent than the surrounding words in an utterance that are not marked with a pitch accent.

In languages such as English and Dutch, the distribution of pitch accents within an utterance tends to be related to the discourse context. More specifically, accents tend to occur on words that represent new information, whereas given information or information that can be derived from the preceding context is typically de-accented. For instance, an utterance like "Mandy bought a book" would get a main accent on either the first or the last word of the sentence, depending on whether it was an answer to (2) or (3), given that these words represent the important new information.

2. *Who bought a book?*
3. *What did Mandy buy?*

But while the patterns described above are typical for Germanic languages, such as Dutch, German and English, many other languages do not exploit accent distribution as consistently for informational purposes or even rely on other linguistic devices to achieve the same goal. This is already obvious when we consider the fact that in many languages the presence or absence of an accent is purely lexically determined and does not depend on discourse factors, such as the given/new distinction. For instance, in Japanese, a word like "sankaku" ('triangle') is obligatorily produced with a pitch accent irrespective of the discourse context, whereas "shikaku" ('square') will never get an accent (and would even sound ungrammatical if it should receive one).

The distributional patterns that can be observed in Dutch and English may also not generalize to languages where the position of an accent is partly determined by syntactic rules. In particular, it has been claimed that there is a difference between what has been termed "plastic" and "non-plastic" languages (Ladd 1996; Vallduví 1992). Plastic languages, including Dutch and English, have a relatively fixed word order but a rather flexible prosodic structure. While words cannot easily be moved between different positions in a sentence, readers are relatively free to accent different words independent of their position within a sentence. In contrast, non-plastic languages, such as Italian, Romanian or Catalan, can more freely change word order within a sentence but are more constrained in the extent to which they can accent words in various sentence positions. Empirical evidence for the alleged prosodic differences between plastic and non-plastic languages has been provided by Cruttenden (1993, 1997), Swerts et al. (2002) and Swerts (2007). In particular, when

comparing Germanic and Romance languages, it has been shown repeatedly that it is quite normal to move accents patterns within noun phrases in the former, while the accent distribution is largely fixed in the latter.

The cross-linguistic facts presented above are important from the perspective of language learning. Obviously, if one is acquiring a new language, it is important not only to learn the words and grammatical aspects of that new language, but also to acquire its rules with respect to accent placement in relation to word order laws. Swerts and Zerbian (2010) examined how the differences in accent distribution between languages can have repercussions for L2 learning. Earlier studies have shown that characteristics of a person's first language (L1) may transfer to a second language (L2); for instance, speakers may produce sentences with an atypical word order, may have difficulties with the gender of nouns or with the application of specific morphosyntactic rules when they express themselves in a non-native language because of interference from their first language. Swerts and Zerbian (2010) looked at the extent to which this holds for aspects of intonation as well. They investigated whether traces of the L1 can be discerned in the way intonation is used in the L2 for two functions: (1) to highlight certain words by making them sound more prominent and (2) to signal continuation or finality in a list by manipulating the speech melody. In particular, the authors explored the way focus and boundaries are marked prosodically in Zulu and also compared such prosodic functions in two variants of English in South Africa, i.e., English spoken as a first language and English[3] spoken as a second/additional language by native speakers of Zulu. Using a specific elicitation procedure, Swerts and Zerbian found that native speakers of South African English mark focused words and position within a list by intonational means, just as in other L1 varieties of English, where Zulu only uses intonation for marking continuity or finality. A second study focused on speakers of Black South African English and compared the prosody of proficient *versus* less proficient speakers. The results of the study show that the production of proficient speakers was perceptually and acoustically equivalent to the output of native speakers of English, both with respect to their use of intonation for marking focus and boundaries. The less proficient speakers, on the other hand, marked boundaries in a similar way as L1 speakers of English but did not use prosody for signaling focus, i.e., they followed the rules of their native language. To sum up, the study by Swerts and Zerbian (2010) gives a clear example of how less proficient L2 learners transfer specific prosodic functions from the L1 to L2.

To sum up, so far we have seen that in spite of some universal tendencies in the implementation and interpretation of prosody, expressed with Gussenhoven's Biological Codes, there are comparative differences in both global and functional prosodic conventions. Moreover, the L1 conventions may transfer to the L2 production of less proficient speakers and thus negatively affect their linguistic performance. The question we will attempt to answer in the following section is how L1 learners acquire prosodic features in the native language. In particular, we will focus on prosodic adaptation (imitation).

[3]This English variety is commonly referred to as Black South African English.

6.3 Prosodic Adaptation in L1 Learning

When we interact, we take over each other's behavior both consciously and automatically. This phenomenon is most clearly observable in interaction with children who mimic the behavior and speech mannerisms of older children and adults (Meltzoff and Moore 1977; Termine and Izard 1988). Even infants no more than a few months old are likely to open their mouth or stick out their tongue if the person they interact with exhibits this kind of playful conduct. This phenomenon is sometimes referred to as the 'chameleon effect' (Chartrand and Bargh 1999) and apart from child-adult interactions, it has been detected in a number of natural settings; for example, students in a class would take over the teacher's postures (Bernieri 1988; LaFrance 1979, 1982; LaFrance and Broadbent 1976) and counselors would take over some behavioral features from their clients (Maurer and Tindall 1983).

Imitation is overwhelmingly present in the area of auditive prosody. Already early on, parents engage in vocal imitations with their infants, possibly to help them become involved in reciprocal vocal play (Papoušek et al. 1987). Moreover, it is well known that child-directed speech typically involves raised pitch, as well as widened pitch range and reduced speech rate (Lieberman 1984; Fernald et al. 1989). These prosodic features are assumed to capture and maintain the child's arousal and attention (Fernald 1989, 1992). According to anecdotal evidence reported by Lieberman (1984), however, babies in their non-cry vocalizations take over the average pitch of their speaking parents, using lower pitch in interactions with fathers and higher pitch when communicating with their mothers. Interestingly, quantitative studies offer mixed evidence both in favor and against the claim. In a study of 9–12 month old children, Siegel et al. (1990) found no indication of pitch imitation either in a laboratory or a home setting. A longitudinal study of a single infant over the period from 3 to 17 months (McRoberts and Best 1997) also did not result in any support for the claim that infants adjust their vocal pitch in the direction of the parents. On the other hand, Papoušek and Papoušek (1989) report a similarity in a slight majority of adjacent utterances produced in child-mother interactions. In their longitudinal study spanning 14 months, they evaluated infant vocalizations in the context of their mothers' preceding and following utterances with respect to a number of prosodic and microprosodic features (absolute pitch, pitch contour shape, duration, rhythm, vowel-like resonance and consonant-like closure), using both an auditory assessment, as well as acoustic measurements. Overall, absolute pitch was the most frequently matched prosodic feature but the data showed considerable individual variability. With respect to other prosodic properties, Guitar and Marchinkoski (2001) reported on speech rate adaptation in semi-natural interactions between mothers and toddlers. In general, it is relevant to note that the experimental studies of prosodic adaptation in early age are difficult because it is impossible to use priming paradigms to control for the difference between non-imitative spontaneous vocalizations and imitative behavior and also to distinguish maternal imitation from infant imitation.

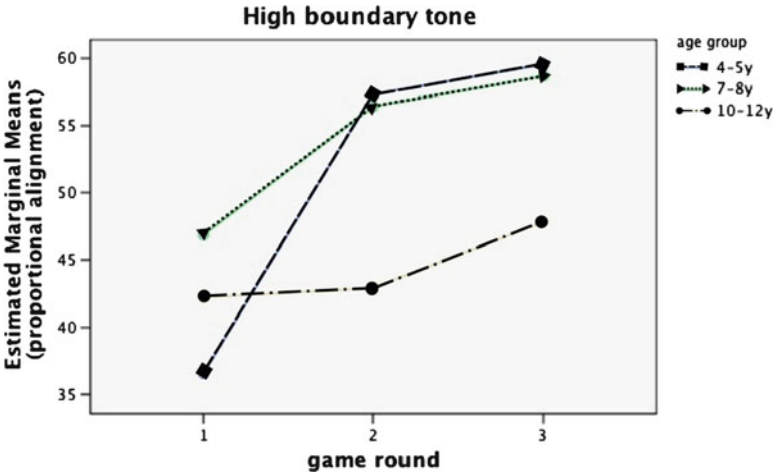


Fig. 6.1 A comparison of high-boundary tone adaptation across different age groups over the course of three game rounds

Studies of older children make use of more constrained experimental settings in which prosodic adaptation can be measured with a higher reliability. Ryalls and Pisoni (1997) observed accommodation in duration by children shadowing an adult voice. The research of Frome Loeb and Allen (1993) with 3- and 5-year-old preschoolers involved a sentence elicitation task in which the children were primed with three different intonational contours (declarative, interrogative or flat). A perceptual and acoustic analysis of the recordings showed an effect of the prime that was, moreover, stronger for the older age group. Coulston et al. (2002) found that children interacting with an extroverted or introverted animated character adapted the amplitude in their speech in accordance with the character's personality, speaking louder to an extroverted avatar and softer to an introverted one. In an experimental study of different age groups (4- to 12-year-olds, all native Dutch speakers), Nilsenová et al. (2009) made use of a game paradigm to study both global and local pitch adaptation, making use of manipulated re-synthesized speech. In the game, participants looked for hidden treasure chests behind playing cards depicting Disney movie characters, animals and toys. In a series of six games, the children were taking turns with the computer to name the cards. They were primed with speech involving either low or high boundary tones (e.g., *Ik neem PLUTO*$_{H\%}$ – "I'll take PLUTO") or with a globally raised or lowered pitch. The data showed that children in all age groups adapted to the boundary tones with which they were primed (with the oldest group adapting the least, *viz.* Fig. 6.1),[4] but did not adapt to the global pitch changes.

[4]Cf. Nilsenová and Nolting (2010) who report a stronger adaptation in pitch span and f_0 at the utterance boundaries (the final f_0 in the prime and initial f_0 in the participant's response) by 4- to 6-year-old children compared to adult speakers.

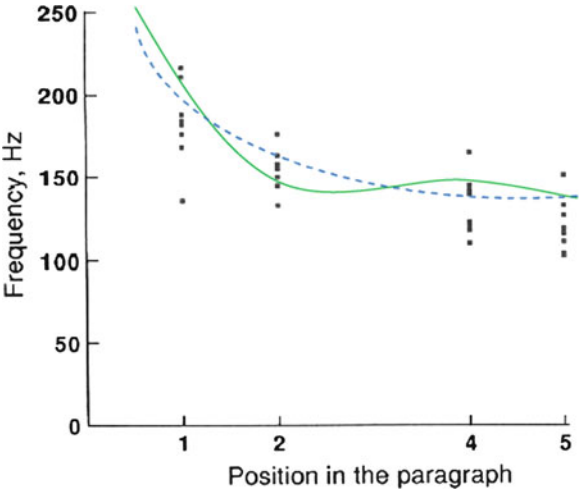


Fig. 6.2 The effect of pitch priming in a game of re-telling stories (*full line*) plotted against the natural effect of the Production Code (*dotted line*, adapted from Sluijter and Terken 1993:184). The speakers were primed with a sequence of high – low – high – low global pitch

In a subsequent study, it was examined if the length of the interaction and the social presence of the computer character possibly had an effect on global pitch accommodation. A group of 62 school children (4–5- and 6–7-year-olds; 32 male) interacted with various cartoon characters in a game of re-telling stories. They first heard a story told to them by a cartoon character with a matching voice that was manipulated and re-synthesized in order to create either a high- or a low-pitched voice. Subsequently, they retold the story to a new character and were consistently rewarded for precision. An acoustic analysis of the recordings showed an adaptation in mean pitch for both age groups and genders. When the cartoon character told a story in a high-pitched voice, the children re-told the story with a higher pitched voice as well. There was also an effect of global declination in that the pitch height was lower with each story round; however, the effect of pitch priming resulted in a lower mean pitch in the conditions with low cartoon voices and a higher mean pitch in the conditions with high cartoon voices (viz. Fig. 6.2).

To sum up, we have seen that from an early age, children tend to imitate the prosodic features of their environment. Moreover, comparative studies of different age groups suggest that the effect decreases with age. This fact implies that prosodic adaptation is especially important in the early stages of L1 learning.

While the tendency to mimic the behavior of our interlocutors appears to be unconscious, interestingly, the process can be affected by the affinity we feel towards them. Our relationship to the mimickers influences the degree to which we take over their posture, mannerisms and other behavior. Experimental participants who were instructed to interact in a cooperative task in which it was important to "get along and work well together" were more likely to take over their collaborator's behavior (Lakin and Chartrand 2003). Some speakers – namely those who are highly aware of people around them – have an innate tendency to monitor and imitate their environment (Cheng and Chartrand 2003; Chartrand and Bargh 1999). Rather than just being a by-product of particular cognitive makeup, imitating someone's behavior

seems to fulfill the function of affecting the mimicked persons relationship towards the mimicker. Mimicking someone's behavior is often used to communicate liking for someone (Bavelas et al. 1986) and, in fact, creates rapport. Speakers who are mimicked by their conversational partners like them more and report a higher degree of harmony in the interaction (Chartrand and Bargh 1999); mimicking others thus has a clear social function in that it reflects the degree of affiliation we feel and want to elicit in other people (Lakin et al. 2003).

6.3.1 Social Effects of Prosodic Adaptation

In certain conversational settings, adult speakers have been found to adjust their vocal features to those of their interlocutor. In a series of studies of both natural and semi-natural conversational speech, Gregory and his colleagues have shown that speakers consistently adapt to dialogue partners they consider to be either socially dominant or otherwise attractive (Gregory 1994; Gregory et al. 1997, 1999; Gregory and Webster 1996). For example, an analysis of long-term averaged spectra of 25 interviews on Larry King Tonight (Gregory and Webster 1996) showed that while in general, the participants of the interview converged their vocal properties over time, the accommodation from Mr. King's side was stronger when he was interacting with more prominent guests and weaker when he was interacting with guests ranked lower on celebrity status. In another study, Gregory and Gallagher (2002) examined 19 debates between presidential candidates during the televised U.S. presidential debates. They showed that an acoustic analysis of the candidate's voice could predict the popular vote results in eight of the cases; roughly speaking, in these cases, the winning candidate would have a more stable, less adaptive vocal profile than the other candidate. Finally, Haas and Gregory (2005) recorded and analyzed semi-natural dialogues between two female speakers and concluded that less attractive women were adapting their global pitch to women perceived as being more attractive by independent judges. These and other acoustic analyses of different types of conversational dyads suggest that pitch adaptation is not a symmetric process but that the speaker that perceives/is perceived as having a lower social status is more likely to adapt to the speaker with a higher social status.

So far, we have focused mainly on pitch adaptation but it has been observed that interlocutors converge on their use of other prosodic features as well. Pardo (2006) reports a general phonetic convergence in a map task dialogue, with instruction givers adapting more than instruction followers. In a recent study, Staum Casasanto et al. (2010) found speech rate accommodation in interactions with a virtual agent, partly influenced by the subject's social tie with the agent. Babel (2009) examined the relation between spectral adaptation in a shadowing task and the participant's racial tie to the speaker; she found that if the speaker was perceived to be racially less similar to the participant, spectral adaptation decreased.

In sum, we have identified prosodic imitation as the process behind prosody acquisition in L1 and argued that the phenomenon is influenced by social relations among speakers and also determines their rapport. The idea that adaptation, while

primarily an automatic cognitive process (Pickering and Garrod 2004) that supports language learning and language processing, can also enhance social mechanisms has been theoretically implemented in various ways. On the one hand, Dijksterhuis and Bargh (2001) proposed that the process is due to the activation of mirror neurons but can be inhibited by social knowledge, e.g., by self-focused attention. Gregory and his collaborators (see also Wichmann 2010), on the other hand, argue that the process is triggered by social knowledge in certain contexts if the speaker likes the interlocutor or wishes to signal a need to belong, as well as by the possible dialogue function of the speaker (instruction giver *vs*. instruction follower). Branigan et al. (2010) suggested that the function of the adaptive process depends on the nature of the mechanism; for instance, it could be that adaptation of gradient prosodic properties such as global pitch or intensity primarily serves social purposes, while the imitation of categorical prosodic features such as boundary tones mainly boosts language processing (Nilsenová et al. 2009). Whatever the theoretical explanation turns out to be, it is obvious that the process forms an integral part of human communication and as such, is crucial to second language learning.

While prosodic adaptation in second language dialogue has so far not received much attention in theoretical and experimental studies, in other contexts, it has been noted that adaptation (alignment) by second language learners is often impaired (Costa et al. 2008). For instance, L2 speakers often fail to adapt to L1 speaker's choice of syntactic constructions or lexical expressions, as in the dialogue in (4) adopted from Costa et al. (2008:538):

4. L2 speaker: I need a piece of paper with nothing on it
   L1 speaker: A blank sheet of paper?
   L2 speaker: Yeah, a blank piece of paper

Failed adaptation may have a negative effect on the automatic retrieval of linguistic representations and on the successful construction of mental models of the situation under discussion. While a number of studies have shown that L1 speakers adapt their speech to L2 speakers (so called 'foreigner talk') by using shorter sentences, higher pitch and slower rate (Snow 1995), we have seen that L2 speakers' adaptation is just as important and may even enhance performance at other levels of linguistic representations (Pickering and Garrod 2004; Costa et al. 2008). In fact, preliminary studies of the effects of dialogue adaptation on learning show that prosodic adaptation may be one of the factors predicting learning in a student-tutor dialogue (Ward and Litman 2007). In the next section, we will discuss pragmatic implications of the results for L2 learning of prosody.

6.4 Tools for Prosody Learning

"Of all the elements of a target language, the intonation appears to be the most difficult to acquire…" (Leon and Martin 1972, as cited by Spaai and Hermes 1993:21). One of the main problems in acquiring L2 prosody is that learners are often not aware of the differences between the L1 and L2. Even when explicitly instructed to

imitate L2 patterns, they are likely to fail in generalizing them to newly generated sentences. Moreover, a purely perception-based observation of the sort "there is a slight fall followed by a rise at the end of questions" is never entirely reliable (Mertens 2004). For example, a final rise can be perceptually confused with a fall if accompanied with a rapid drop in intensity. The impression of a rising tune can be due to various factors, such as a high boundary tone, a high nuclear pitch accent on the stressed syllable (with a subsequent final fall), a lack of final declination or an overall high pitch of the utterance compared to what is assumed to be the average pitch of the speaker. The impression of high tempo can be due to the number of melodic movements in the L2, as in the comparison between Italian and Dutch above. These prosodic features may carry a certain semantic import (e.g., a high boundary tone conveying lack of speaker's commitment, a high nuclear pitch accent marking "new" information, lack of final declination signaling incompleteness and overall high pitch being typical for questions) and it is thus important that L2 learners receive reliable information about their occurrence.

Contrary to pure auditory analysis, instrumental analysis of prosodic features, where acoustic properties of the speech signal are visualized with computer software, can provide the needed information. Speech-visualizing technology has been employed in the teaching of intonation for a number of years, starting with early efforts in the 1960s (Levis and Pickering 2004). Since then, both software and hardware systems have become available to assist in L2 intonation learning (Spaai and Hermes 1993; Moltholt 1988; Boersma and Weenink 2012). Originally, participants were simply instructed to imitate utterances produced by a native speaker, trying to match the original pitch contour as closely as possible. Given that intonation is largely context-dependent, it is, in fact, necessary to abandon the sentence-level practice in favor of discourse-level based approaches (Levis and Pickering 2004); preferably, L2 learners should be confronted with dialogue situations involving two speakers in interaction. It is important that they are aware of the functions of prosodic adaptation and are taught – both explicitly and implicitly (e.g., by engaging them in collaborative language games with an L1 speaker) to employ prosodic adaptation in their language production. Observing imitative behavior in a dialogue also appears to be beneficial to learning. In particular, Parrill and Kimbara (2006) showed that when observers are asked to reproduce a conversation, involving speech (or gesture) mimicry, they are more likely to use the mimicked features as well.

6.5 Conclusion

We have shown that while the use of prosody is partly based on universal principles, the global implementation of prosodic features and their specific functions are largely language dependent. Therefore, language learners need to receive explicit instructions regarding the differences between their L1 and L2. We argued that prosodic adaptation is crucial to the process of acquiring L2 prosody from a number of perspectives: (1) it appears to be the natural way in which children pick up the

prosody of their mother tongue, (2) at least for L1 speakers, it presumably supports language processing on a number of representational levels, (3) its presence is positively correlated with learning in student-tutor dialogues, and (4) it is a powerful device employed by speakers to signal social relations in a conversation. In order to become fully aware of the phenomenon of prosodic adaptation, L2 learners can make use of widely available speech-visualizing computer tools; these tools can best be employed in combination with contextually situated dialogues in which L1 speakers can be observed adapting to each other's speech properties.

References

Atterer, M., and D.R. Ladd. 2004. On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics* 32: 177–197.

Babel, M. 2009. Phonetic and social selectivity in speech accommodation. Unpublished PhD dissertation, University of California, Berkeley.

Bavelas, J.B., A. Black, C.R. Lemery, S. Macinnis, and J. Mullet. 1986. Experimental methods for studying "elementary motor mimicry". *Journal of Nonverbal Behavior* 10: 102–119.

Bernieri, F.J. 1988. Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior* 12: 120–138.

Boersma, P., and D. Weenink. 2012. Praat: Doing phonetics by computer [Computer program]. Version 5.3.04. Retrieved 12 Jan 2012 from http://www.praat.org/.

Bolinger, D. 1978. Intonation across languages. In *Universals of human language. Vol II: Phonology*, ed. J. Greenberg, 471–524. Palo Alto, CA: Stanford University Press.

Branigan, H.P., M.J. Pickering, J. Pearson, and J.F. McLean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42: 2355–2368.

Brent, M.R., and J.M. Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81: B33–B44.

Bruce, G., and I. Thelander. 2001. A pitch accent journey in southern Sweden. *Working Papers* 49, 14–17, Department of Linguistics, Lund University.

Chartrand, T.L., and J.A. Bargh. 1999. The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76: 893–910.

Chen, A. 2004. Universal and language-specific perception of paralinguistic intonational meaning. PhD thesis, LOT dissertation series, 102.

Cheng, C.M., and T.L. Chartrand. 2003. Self-monitoring without awareness: Using mimicry as a nonconscious affiliation strategy. *Journal of Personality and Social Psychology* 86: 1170–1179.

Collier, R. 1991. Multi-language intonation synthesis. *Journal of Phonetics* 19: 61–73.

Costa, A., M.J. Pickering, and A. Sorace. 2008. Alignment in second language dialogue. *Language & Cognitive Processes* 23(4): 528–556.

Coulston, R., S. Oviatt, and C. Darves. 2002. Amplitude convergence in children's conversational speech with animated personas. In *Proceedings of the 7th International Conference on Spoken Language Processing*, Denver, 2689–2692.

Cruttenden, A. 1993. The de-accenting and re-accenting of repeated lexical items. In *Proceedings of the ESCA Workshop on Prosody*, Lund, Sweden, Sept 1993.

Cruttenden, A. 1997. *Intonation*. Cambridge: Cambridge University Press.

den Os, E. 1988. Rhythm and tempo of Dutch and Italian: A contrastive study. PhD thesis, Utrecht University, Utrecht, the Netherlands.

Dijksterhuis, A., and J.A. Bargh. 2001. The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Psychology* 33: 1–40.

Fernald, A. 1989. Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development* 60: 1497–1510.

Fernald, A. 1992. Human vocalizations to infants as biologically relevant signals: An evolutionary perspective. In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*, ed. J.H. Barkow, L. Cosmides, and J. Tooby, 365–388. Hillsdale: Lawrence Erlbaum Associates.

Fernald, A., T. Taeschner, J. Dunn, M. Papoušek, B. Boysson-Bardies, and I. Fukui. 1989. A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language* 16: 477–501.

Frome Loeb, D., and G.D. Allen. 1993. Preschoolers' imitation of intonation contours. *Journal of Speech and Hearing Research* 36: 4–13.

Gregory, S.W. 1994. Sounds of power and deference: Acoustic analysis of macro social constraints on micro interactions. *Social Perspectives* 37: 497–526.

Gregory, S.W., and S. Webster. 1996. A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology* 70: 1231–1240.

Gregory, S.W., K. Dagan, and S. Webster. 1997. Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior* 21: 23–43.

Gregory, S.W., T.J. Gallagher, and B.E. Green. 1999. Spectral analysis of candidates nonverbal communication predicts national debate outcomes. Paper presented to the American Sociological Association meetings, Chicago, IL.

Gregory, S.W., Jr. and T.J. Gallagher. 2002. Spectral analysis of candidates' nonverbal vocal communication: Predicting U.S. presidential election outcomes. *Social Psychology Quarterly* 65: 298–308.

Grice, M., M. D'Imperio, M. Savino, and C. Avesani. 2005. Strategies for intonation labeling across varieties of Italian. In *Prosodic typology: The phonology of intonation and phrasing*, ed. S.-A. Jun, 362–389. Oxford: Oxford University Press.

Guitar, B., and L. Marchinkoski. 2001. Influence of mothers' slower speech on their children's speech rate. *Journal of Speech, Language, and Hearing Research* 44: 853–861.

Gussenhoven, C. 2002. Intonation and interpretation: Phonetics and phonology. In *Proceedings of speech and prosody'02*, Aix-en-Provence, ed. B. Bel and I. Marlien.

Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Haas, A., and S.W. Gregory. 2005. The impact of physical attractiveness on women's social status and interactional power. *Sociological Forum* 20: 449–471.

Halliday, M.A.K. 1967. Notes on transitivity and theme in English. Part 2. *Journal of Linguistics* 3: 199–244.

Hart, J.T., R. Collier, and A.A. Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Irino, T., and R.D. Patterson. 2002. Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-mellin transform. *Speech Communication* 36: 181–203.

Juslin, P.N., and P. Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129(5): 770–814.

Kunzel, H. 1989. How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46: 117–125.

Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

LaFrance, M. 1979. Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. *Psychology Quarterly* 42: 66–70.

LaFrance, M. 1982. Posture mirroring and rapport. In *Interaction rhythms: Periodicity in communicative behavior*, ed. M. Davis, 279–298. New York: Human Sciences Press.

LaFrance, M., and M. Broadbent. 1976. Group rapport: Posture sharing as a nonverbal indicator. *Group and Organization Studies* 1: 328–333.

Lakin, J.L., and T.L. Chartrand. 2003. Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science* 14: 334–339.

Lakin, J.L., V.E. Jefferis, C.M. Cheng, and T.L. Chartrand. 2003. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27: 145–161.

Leon, P.R., and P. Martin. 1972. Applied linguistics and the teaching of intonation. *The Modern Language Journal* 56: 139–144.

Levis, J., and L. Pickering. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32: 505–524.

Lieberman, P. 1984. *The biology and evolution of language*. Cambridge: Harvard University Press.

Liberman, M., and J. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In *Langauge sound structure*, ed. M. Aronoff and R. Oehrle, 157–233. MIT Press, Cambridge MA.

Maurer, R.E., and J.H. Tindall. 1983. Effect of postural congruence on client's perception of counselor empathy. *Journal of Counseling Psychology* 30: 158–163.

McRoberts, G.W., and C.T. Best. 1997. Accommodation in mean f_0 during mother-infant and father-infant vocal interactions: A longitudinal case study. *Journal of Child Language* 24: 719–736.

Meltzoff, A.N., and M.K. Moore. 1977. Imitation of facial and manual gestures by human neonates. *Science* 198: 75–78.

Mertens, P. 2004. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model, In *Proceedings of speech prosody 2004*, Nara, Japan, 23–26.

Moltholt, G. 1988. Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly* 22: 91–111.

Nilsenová, M., and P. Nolting. 2010. Linguistic adaptation in semi-natural dialogues: Age comparison. In *Text, speech and dialogue*, ed. P. Sojka, A. Horák, I. Kopeček, and K. Pala. Berlin/ Heidelberg: Springer.

Nilsenová, M., M. Swerts, V. Houtepen, and H. Dittrich. (2009) Pitch adaptation in different age groups: Boundary tones versus global pitch. In *Proceedings of Interspeech 2009: 10th Annual Conference of the International Speech Communication Association*, ed. M. Uther, R. Moore, and S. Cox, 1015–1018. Brighton: ISCA.

Ohala, J. 1983. Cross-language use of pitch: An ethological view. *Phonetica* 40: 1–18.

Ohala, J. 1984. An ethological perspective on common cross-language utilization of F0 in voice. *Phonetica* 41: 1–16.

Papoušek, M., H. Papoušek, and B.J. Harris. 1987. The emergence of play in parent-infant interactions. In *Curiosity, imagination, and play: On the development of spontaneous cognitive and motivational processes*, ed. D. Görlitz and J.F. Wohlwill. Hillsdale: Erlbaum.

Papoušek, M., and H. Papoušek. 1989. Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language* 9: 137–157.

Pardo, J. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119(4): 2382–2393.

Parrill, F., and I. Kimbara. 2006. Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior* 30: 157–166.

Pickering, M., and S. Garrod. 2004. Towards a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences* 27: 169–190.

Rietveld, A., and C. Gussenhoven. 1987. Perceived speech rate and intonation. *Journal of Phonetics* 15: 273–285.

Ryalls, B.O., and D.B. Pisoni. 1997. The effect of talker variability on word recognition in preschool children. *Developmental Psychology* 33: 441–452.

Scherer, K. 1989. Vocal correlates of emotional arousal and affective disturbance. In *Handbook of social psychophysiology*, ed. H. Wagner and A. Manstead, 165–197. New York: Wiley.

Siegel, G.M., M. Cooper, J.L. Morgan, and R. Brenneise-Sarshad. 1990. Imitation of intonation by infants. *Journal of Speech and Hearing Research* 33: 9–15.

Snow, C. 1995. Issues in the study of input: Fine-tuning, universality, individual and developmental differences and necessary causes. In *The handbook of child language*, ed. P. Fletcher and B. MacWhinney, 180–194. Oxford: Blackwell.

Sluijter, A.M.C., and J.M.B. Terken. 1993. Beyond sentence prosody: paragraph intonation in Dutch. *Phonetica* 50: 180–188.

Soderstrom, M., M. Blossom, R. Foygel, and J.L. Morgan. 2008. Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language* 35: 869–902.

Spaai, G.W.G., and D.J. Hermes. 1993. A visual display for the teaching of intonation. *CALICO Journal* 10: 19–30.

Staum Casasanto, L., K. Jasmin, and D. Casasanto. 2010. Virtually accommodating: Speech rate accommodation to a virtual interlocutor. In *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, ed. S. Ohlsson and R. Catrambone, 127–132. Austin: Cognitive Science Society.

Swerts, M. 2007. Contrast and accent in Dutch and Romanian. *Journal of Phonetics* 35: 380–397.

Swerts, M., and S. Zerbian. 2010. Intonational differences between L1 and L2 English in South Africa. *Phonetica* 67: 127–146.

Swerts, M., E. Krahmer, and C. Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30: 629–654.

Termine, N.T., and C.E. Izard. 1988. Infants' response to their mothers' expressions of joy and sadness. *Developmental Psychology* 24: 223–229.

Vallduví, E. 1992. *The information component*. New York: Garland Press.

Verhoeven, J.De, G. Pauw, and H. Kloots. 2004. Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech* 47: 297–308.

von Kriegstein, K., J.D. Warren, D.T. Ives, R.D. Patterson, and T.D. Griffiths. 2006. Processing the acoustic effect of size in speech sounds. *NeuroImage* 32: 368–375.

Ward, A., and D. Litman. 2007. Dialog convergence and learning. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work*, Los Angeles, CA, 262–269.

Wichmann, A. 2010. Intonational meaning in institutional settings: The role of syntagmatic relations. *Cultural Studies of Science Education* 5: 849–857.

Willems, N.J., R. Collier, and J. 't Hart. 1988. A synthesis scheme for British English intonation. *Journal of the Acoustical Society of America* 84: 1250–1261.

Part II
Pragmatics, Prosody and Communication

Chapter 7
Prosody and Meaning: Theory and Practice

Tim Wharton

7.1 Introduction

A recently published encyclopaedia of pragmatics (Cummings 2009) contains over 300 entries. A myriad of different perspectives are adopted, each clustering around a central notion of pragmatics as the study of 'language in use'. There are 'clinical' and 'computational' perspectives; 'developmental', 'experimental', 'formal', 'intercultural', 'optimality-theoretic' perspectives and many more. As well as this, pragmatics is applied to a range of communicative phenomena: the pragmatics of translation, of word-learning, of writing, of communication aids, even (perhaps surprisingly) of non-human animal communication. This diversity is at least partially reflected in the range of approaches to pragmatics adopted in this volume: pragmatics is clearly many things to many people.

The conception of pragmatics adopted in this chapter owes a great deal to the man who was – arguably – the father of modern pragmatics. For Paul Grice the meaning a speaker intended to convey was to be understood in terms of propositional attitude psychology. Ultimately, the meaning of utterances (or sentences, or words) was to be characterised in terms of the beliefs, desires and intentions of the people who utter them. Beliefs, desires and intentions are all psychological/cognitive phenomena and so this chapter takes a psychological/cognitive[1] view of pragmatics, regarding it as – ultimately at least – a branch of psychology or cognitive science.

No-one denies that the way we say the words we say makes a substantial contribution to our intended meaning. In the most intuitively obvious cases, a particular tone of voice might indicate that we want to dissociate ourselves entirely from the

[1]The small 'c' is important.

T. Wharton (✉)
Department of Media and Communication, Kingston University, London, UK
e-mail: t.wharton@kingston.ac.uk

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_7,

proposition we are expressing: that we *mean* the opposite of what we are saying. So it is also true that the way we say what we say is capable of conferring on those words entirely new layers of meaning. But prosody also works in other, more subtle ways. In English, the way words are grouped together into intonation phrases conveys information about constituency relations and grammatical structure. Within these phrases, differences in the volume, length and pitch of syllables help to direct a listener's attention to the most salient points of a message.

However, until relatively recently, most of the investigation into the complex interaction between prosody and speaker meaning has been done not by pragmatists but by phonologists (Halliday 1963, 1967; O'Connor and Arnold 1973; Brazil 1975; Ladd 1978, 1996; Bolinger 1983a, b; Ward and Hirschberg 1988; Hirschberg and Ward 1995; Pierrehumbert and Hirschberg 1990; Gussenhoven 1984, 2002, 2006; Chen and Gussenhoven 2003). The various schools and traditions which have emerged have made great progress and we now have formal analyses of prosodic structures and systems in a number of different languages and concrete proposals on how they relate to meaning.

While this body of work often appeals – to take one example – to 'systems of rich interpretive pragmatics' (Ladd 1996, p. 39), very little utilises a recognised theoretical pragmatic framework, particularly the kind of framework that takes Grice's views on meaning – or indeed a cognitive view of pragmatics – seriously. It is not enough to write – as D. Robert Ladd did in 1996 when attempting to assess the merits of competing accounts of intonational meaning – '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' (Ladd 1996, p. 101). We know much more about pragmatic inference than we did 15 years ago. Recent work in cognitive sciences on reasoning and rationality (Gigerenzer et al. 1999; Gigerenzer 2000) has developed work on 'bounded rationality' and proposes that evolution has left humans with rules-of-thumb – heuristics – which enable us to use what is, after all, a finite cognitive capacity in the most economical way. Work within cognitively oriented approaches to pragmatics (Blakemore 2002; Carston 2002; Sperber and Wilson 1986/1995) makes concrete proposals about these fast, frugal pragmatic heuristics. As well as this, since the early 1990s researchers working within pragmatics have been looking at the kind of issues raised by linguists and phonologists during the previous 25 years and there is now a rich literature that considers prosodic contributions to meaning from a cognitive pragmatics-based perspective rather than a phonological one (Clark 2007; Clark and Lindsey 1990; Escandell-Vidal 1998; Fretheim 2002; House 1990, 2006; Imai 1998; Vandepitte 1989; Wichmann 2002; Wilson and Wharton 2006). Wichmann and Blakemore's (2006) volume is the result of a conference in 2002, which was a bold attempt to bring people from the two disciplines together.

The aims of this chapter are two-fold. Firstly, I intend to give the reader a flavour of some of the various dimensions along which the debates on the relationship between pragmatics (as conceived in this chapter) and prosody take place, as well as suggest ways in which studies in this area might be advanced. In this

regard, I focus on three theoretical questions (these questions are explored in more detail in Wharton 2009, in press):

1. How can the different types of prosody be characterised?
2. What is the relationship between prosody and *intentional* communication?
3. What kind of meaning does prosody encode (if anything)?

The next section of the chapter is divided into three sub-sections, each of which deal with the questions above (so question (1) is dealt with in Sect. 7.2.1, question (2) in Sect. 7.2.2 and question (3) in Sect. 7.2.3).

Secondly, I would like (as my title suggests) to extend the debate to the practical domain. To what extent is the theoretical debate reflected in the teaching of English pronunciation? Can the theory usefully inform the practice?

7.2 Theory

7.2.1 How Can the Different Types of Prosody Be Characterised?

It is clear that the various elements of prosody do not all work in the same way. On the one hand, the lexical distinction in English between the verb and the noun 'export' or between the preposition 'below' and the verb 'billow' are seen as properly linguistic, as are the tonal lexical contrasts in languages such as Burmese and Thai. On the other, affective prosody, say, sounding surprised or bored, or happy or sad, are natural phenomena (Ladd 1996; Gussenhoven 2004) interpreted by non-linguistic systems. Grice himself was sensitive to this distinction:

> We might start by trying to think of stress as a purely natural way of highlighting, or making prominent a particular word: compare putting some object (e.g., a new hat) in an obvious place in a room so that someone coming into the room will notice or pay attention to it. But there are various suggestible ways of doing this with a word: e.g., intoning it, saying it in a squeaky voice. Such methods would not be thought just unusual, but would be frowned upon […] So there is a good case for regarding stress as a conventional device for highlighting.
>
> Grice (1967, Lecture III, pp. 17–18)

Halliday's (1963, 1967) account of prosody was based within a broad tradition which regarded prosody as properly linguistic. He aimed to articulate a theory of grammar that was rich enough to accommodate intonation patterns and aimed to extend the notion of language to incorporate all prosody: in other words, he was working toward a *semantic*, rather than pragmatic explanation. Other linguistically-oriented accounts of prosody can be found in the works of Sag and Liberman (1975) and Gussenhoven (1984).

In direct contrast to these accounts, Bolinger (1983a) was famously of the view that we would be better to focus more on the natural, pragmatic side. He focused on

the interaction between intonation and other, parallel natural components of the complex communicative stimulus:

> If intonation is part of a gestural complex whose primitive and still surviving function is—however elaborated and refined—the signalling of emotions and their degrees of intensity, then there should be many obvious ways in which visible and audible gesture are coupled to produce similar and reinforcing effects. This kind of working parallel is easiest to demonstrate with exclamations. An ah! of surprise, with a high fall in pitch, is paralleled by a high fall on the part of the eyebrows … A similar coupling of pitch and head movement can be seen in the normal production of a conciliatory and acquiescent utterance such as "I will" with the accent at the lowest pitch—we call this a bow when it involves the head, but the intonation bows at the same time. (1983a, p. 98)

However, he stressed that behaviours may indeed be 'more' or 'less' natural, implicitly suggesting they exist along some kind of continuum and argued that although we may feel some aspects of intonation to be properly linguistic, they still have their roots in the natural behaviours from which they evolved:

> Intonation … assists grammar – in some instances may be indispensable to it – but it is not ultimately grammatical … If here and there it has entered the realm of the arbitrary, it has taken the precaution of blazing a trail back to where it came from (1983b, pp. 106–108).

More recently, it has been proposed that the differences between the various aspects of prosody might be captured by suggesting that prosodic effects range along a continuum from 'natural' to language-specific (Gussenhoven 2002; Pell 2002). According to various phonologists, prosody encodes both *linguistic* and *paralinguistic* meaning.

Turning first to the 'natural' end of the natural/language-specific continuum, Chen and Gussenhoven (2003) argue that the interpretation of paralanguage is governed by *biological codes*. An example of one such code is the *effort code*, which links the amount of energy expended in the production of speech to a range of interpretive effects. An increase in effort may, for example, lead to increased articulatory precision, creating an impression of 'helpfulness', or 'obligingness'; or it may result in a wider pitch range, creating an impression of 'forcefulness' or 'certainty' or conveying affective meanings such as 'agitation' or 'surprise'. There are two issues I would like to raise here.

The first of these concerns the notion of *code*. One observation of previous characterisations of the natural aspects of prosody is that all such aspects are analysed as codes. Much work on human communication makes this assumption and adopts a code model of communication, according to which a communicator's thoughts are translated into a signal by the use of a code and translated back from the signal by an audience into the original message. I will argue that biological communicative systems (as distinct from a linguistic prosodic code) do exist, but not quite in the way Gussenhoven envisages: many natural aspects of prosody, however, are not *codes* at all. Marc Hauser (1996) applies a distinction between signs and signals to cases of information transmission among animals. *Signs* carry information by providing evidence for it. *Signals*, on the other hand, are those behaviours that convey information and have been "moulded by natural selection to do so" (Seeley 1989, p. 547). Put differently, the adaptive function of a behaviour is the effect which is historically

responsible for the reproduction and propagation of that behaviour within a species (Millikan 1984; Origgi and Sperber 2000; Sperber 2007).

Whilst a sign may happen to carry information for an observer, it would go on being produced whether or not it carried this information. The presence of chimpanzee nests in trees indicates that chimps live in the area. Certain species of prey – such as forest monkeys – might use the presence of nests to detect whether not its chief predator is nearby (see Hauser 1996, pp. 9–10). The nests, however, cannot be said to have a signalling function.

One way of describing this is to say that natural signs (e.g., the nests) are not *inherently* communicative. They are, in fact, classic cases of natural meaning (meaning$_{N}$) in the Gricean (1957, 1989) sense – see (4), (5), (6). These can be contrasted with Gricean non-natural meaning (meaning$_{NN}$) – see (7) and (8):

4. Those black clouds mean$_{N}$ rain (see Grice 1989, p. 291).
5. Those chimpanzee nests mean$_{N}$ chimps live here.
6. That paw-print means$_{N}$ a bear has passed.
7. That remark means$_{NN}$ 'it's going to rain'.
8. 'Det kommer til å regne' means$_{NN}$ 'it's going to rain'.

By contrast, signals have a communicative function. The function of, say, the honeybee's dance is to inform other honeybees about the location of nectar; the function of the bull-frog's call is to alert female frogs to the fact that he is in the vicinity and looking for a mate. If these behaviours did not carry this information, it would be hard to see why they survive. Most animal communication seems to be based on signalling systems of this type. Since there is no evidence that either honeybees or frogs are capable of multi-layered intention expression and attribution required for their communicative behaviours to be characterised as cases of meaning$_{NN}$, it is hard to see how a system so complex is governed by anything but an innately determined code.

In Wharton (2003b*;* 2009, in press) I illustrate the distinction between natural signs and signals in the human case by comparing shivering with smiling. Shivering is a natural behaviour whose function is to generate heat by rapid muscle movement. It may provide evidence to an observer that an individual is feeling cold. However, it is not its function to carry this information: it is not a signal but a sign. Smiling, by contrast, appears to have evolved as a signalling activity whose function *is* to convey information to others (van Hooff 1972; Ekman 1989, 1992, 1999; Fridlund 1994). Like the bee dance, the bull-frog calls, the alarm calls of vervet monkeys and a whole range of non-human animal communication systems, they are signals rather than signs.

It is not hard to think of prosodic counterparts to shivering and smiling. A speaker's mental or physical state may affect the prosodic properties of her utterance, enabling a hearer with the appropriate background knowledge or experience to infer whether she is healthy or ill, tired or alert, drunk or sober, etc. As with shivering, these prosodic properties carry information about the speaker's mental or physical state but it is not their function to do so: they are natural signs, interpreted by inference rather than decoding. On the other hand, affective tones of

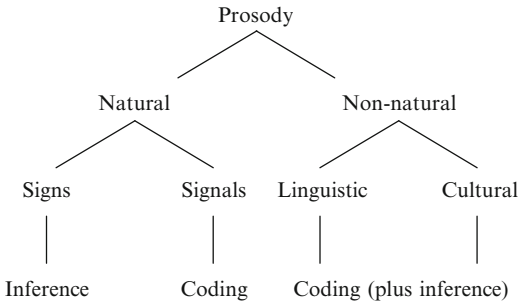


Fig. 7.1 Prosodic inputs to comprehension

voice, like affective facial expressions, may well be natural signals, interpreted by innately determined codes.

Honeybees and frogs both lack the ability to infer the intentions of others but they can still inform each other by means of their dance-based or vocal code. As Grice showed, communication among humans is different. Human linguistic communication exploits the ability to understand the behaviour of others in terms of the intentions behind it – sometimes known as the 'mindreading' ability. A speaker produces linguistically coded evidence of her intention to convey a certain meaning and the hearer must use this as a starting point from which to infer that intention: linguistic communication therefore involves *both* coding *and* inferential intention recognition. And communication among humans not only requires the capacity for inferential intention recognition, but may also be achieved in the absence of any code at all – such as when I nudge my empty glass toward you and you infer that I'd like some more wine.

To return to question (1), I suggest that there are three distinct types of prosodic input: natural signs, which are interpreted purely inferentially; natural signals, which in some cases are interpreted purely by decoding; and *linguistic signals*, which are part of a *linguistic* signalling system, governed by a *linguistic* code with its own special-purpose principles or mechanisms, and interpreted by a combination of decoding and inference. This position is represented in Fig. 7.1 (I turn to discussion of the right-most node in Sect. 7.3):

With the above distinctions in mind, is clear articulation a natural signal interpreted – as Gussenhoven suggests – by an innately determined biological code? I would suggest that it is better analysed as a natural *sign* of the speaker's desire to help the speaker understand, and interpreted via inference and not decoding. A great deal can be communicated by using clear articulation – that you intend to convey helpfulness, or that you are being obliging, or that you want to convey any one of a wide range of other impressions – but nothing is *encoded* by it at all.

The second issue concerns the term *paralanguage*, which – like the term pragmatics – is open to different interpretations (see Wharton 2009, pp. 6–8 for further discussion). For some, paralanguage is defined as including only those vocal aspects of language use that are not strictly speaking part of language. Construed in this way, facial expression, manual and vocal gestures and other kinesic behaviours are *not* part of paralanguage. Yet there are others who take paralinguistic to

include most or all of those aspects of linguistic communication that are not part of language *per se*, but are nonetheless somehow involved with the message or meaning a communicator conveys. The second construal comes closer to how I would want to define paralanguage; rising pitch is so often linked with rising eyebrows that it's perhaps not clear why we would want to say that while the former is part of a paralanguage, the latter is not. Recall Bolinger's words:

> A similar coupling of pitch and head movement can be seen in the normal production of a conciliatory and acquiescent utterance such as "I will" with the accent at the lowest pitch—*we call this a bow when it involves the head, but the intonation bows at the same time*. [My emphasis, TW] (1983a, p. 98)

But if this is the case, where does language end and paralanguage begin? At the end of a careful attempt to motivate the distinction, Ladd (1996, p. 283) concludes: '… I concede that we must stop short of drawing a clear boundary between language and paralanguage. For now that question remains open.' Ladd's fine-grained autosegmental analyses of intonational phonology shed considerable light on which parts of prosody are universal and which are language-specific, but if the distinction between language and paralanguage cannot really be sustained, it is hardly a helpful one.

7.2.2 What Is the Relationship Between Prosody and Intentional Communication?

Natural prosodic signals have a communicative function but neither they nor, of course, natural signs, are intrinsically linked to *intentional* communication. Nonetheless, the sign-signal distinction must be seen in light of the fact that humans have highly-developed metapsychological abilities and natural behaviours, which can therefore be recruited for use in intentional communication. How this takes place is an under-explored domain in studies on pragmatics and prosody.

Sometimes, it may be clear to the audience that an aspect of prosody is being accidentally revealed rather than intentionally conveyed. The speaker's faltering tone of voice may betray the fact that she is a little anxious, or – depending on the context – even that her frustration is boiling over into anger. In more sophisticated cases, the speaker's tone of voice may be covertly manipulated to suggest to an audience that she is accidentally betraying her feelings rather than wanting them to be recognised as part of her meaning in the full Gricean sense. As well as being used *covertly*, a communicator may also *overtly* show her feelings to an audience. She may do this by deliberately producing, and perhaps exaggerating, a natural sign or signal (e.g., a smile or a particular tone of voice); or she may do it by making no attempt to conceal a spontaneously-produced natural sign or signal in circumstances where it is obvious to both communicator and audience that she could have taken steps to conceal them. Grice saw an important difference between this kind of deliberate showing and cases of meaning$_{NN}$ and took pains to distinguish between them. While he was prepared to treat the *simulation* of a piece of natural

behaviour (a frown) as meaning$_{NN}$ and saw its interpretation as crucially involving a process of inferential intention recognition, he argued that a spontaneous piece of natural behaviour, even if openly *shown* to an audience, did not amount to a case of meaning$_{NN}$ (Grice 1989, p. 219).

Grice's distinction has had important consequences for how linguistics and philosophers have come to conceive the domain of pragmatics. To deny that the open showing of spontaneously-produced natural behaviours is a case of meaning$_{NN}$ has had the effect of excluding it from the domain of pragmatics. This may indeed be one of the reasons why those working within pragmatics have tended to gloss over the prosodic aspects of communication. But there seem to be clear cases where the overt showing of spontaneously produced natural signs and signals makes a difference to the speaker's meaning. Consider the utterance (9), in which – given the expletive – we can assume the speaker was making no attempt to conceal the spontaneous anger in her tone of voice (and facial expression):

9. **Damn**! They're up to the roof aren't they? (KCT 190 – British National Corpus)

She would naturally be understood as meaning not only that 'they are up on the roof' but also that she was *angry that they were up on the roof*. Implicatures may depend on this: the degree of anger the speaker overtly shows might warrant the hearer inferring that the speaker is going to take a particular course of action against the people to whom she is referring (give them a slight dressing down or fire them). Consider an utterance of (10), in which a speaker makes no attempt to conceal the emotion in her tone of voice.

10. I am disappointed! (J12 3028 British National Corpus)

The natural tone of voice that the speaker shows to the hearer will not only help him establish the implicit content of her utterance, but will also contribute to the proposition he takes her to be expressing. The truth conditions of her utterance of (10) will vary according to the type or degree of 'disappointed' she intends to communicate (it is a degree term) and hence, reflects in her natural behaviour.

In Wharton (2009) I present a detailed defence of an approach that argues that the open showing of spontaneously-produced natural signs and signals may be located along a continuum between showing and non-natural meaning. My research has a range of implications for both what we take the domain of pragmatics to be and suggests a range of test cases which might be used to better understand the prosodic difficulties found in, for example, people with autism, who typically exhibit impairments in both production and comprehension (Chevallier et al. (2011) is an attempt to explore the extent to which atypical recognition of vocal cues is caused by impaired Theory of Mind).

The first type of test case would consist of natural prosodic *signals* which are *not* overtly shown, and which would not normally be understood as contributing to a communicator's meaning. Examples might be someone trying to conceal their anxiety while speaking, sighing to herself while working alone in her room, or uttering

an exclamation of surprise when something falls off a shelf while no-one else is present. Comprehension of these behaviours in people with prosodic impairments might be compared with comprehension of cases where the same natural prosodic signal is overtly shown in addressing someone, and would normally be understood as contributing to the communicator's meaning.

The second type of test case would consist in natural prosodic *signs* which are not overtly shown and which would not normally be understood as contributing to a communicator's meaning. Examples might be saying "The taxi's arriving" while sounding bored, tired, shaky or ill. Interpretation of these natural prosodic clues might be compared with comprehension of cases where the same natural prosodic sign is overtly shown and would normally be understood as contributing to the communicator's meaning.

The inspiration for the approach developed above comes from *relevance theory* (Sperber and Wilson 1986/1995; Blakemore 2002; Carston 2002). Relevance theorists have consistently argued that there is a continuum of cases between showing and meaning$_{NN}$, *all of which* may fall within the domain of pragmatics and contribute to a speaker's meaning (Sperber and Wilson 1986/1995). Since aspects of the analysis to follow rely on a little background on relevance theory, a brief exposition is in order.

According to the theory, utterance interpretation is a two-phase process. The linguistically encoded logical form, which is the output of the mental grammar, is simply a starting point for rich inferential processes guided by the expectation that speakers will conform to certain standards of communication. In (highly) intuitive terms, an audience faced with a piece of overtly communicative behaviour is entitled to assume that the communicator has a good reason for producing this particular stimulus as evidence not only of their intention to communicate, but of *what* they want to communicate. Thus far, there is no divergence from other post-Gricean and neo-Gricean accounts.

But relevance theory takes the psychology seriously and aims to provide an account of *how* pragmatic inference works. It follows recent work in cognitive science, which sees the mind as an 'adaptive toolbox', a set of dedicated cognitive mechanisms, which have evolved in small steps towards greater *cognitive efficiency* (Hirschfeld and Gelman 1994; Barkow et al. 1995; Sperber 2002):

> Cognitive efficiency involves making the right choices in selecting which available new information to attend to and which available past information to process it with. The right choices in this respect consist in bringing together inputs and memory information, the joint processing of which will provide as much cognitive effect as possible for as little effort as possible. (Sperber 1996, p. 114)

Seen in this way, cognition and communication relies partly on 'fast and frugal heuristics' (Gigerenzer et al. 1999; Gigerenzer 2000), which make it possible to pick out potentially relevant inputs to cognitive processes (e.g., sights, sounds, utterances, memories, conclusions of inferences) and process them in a way that enhances their relevance. Gigerenzer, Todd et al. could be describing one of the fundamental assumptions of relevance theory when they write (p. 21): 'Cognition is the art of focusing on the relevant and deliberately ignoring the rest.'

Construed in this way, the human cognitive system is geared to look out for relevant information, which will interact with information that is already mentally-represented and bring about positive *cognitive effects* based on a combination of the new and the old information. Relevance itself is a property of inputs to cognitive processes and is defined in terms of cognitive effects gained and processing effort expended: other things being equal, the more cognitive effects gained and the less processing effort expended in gaining those effects, the greater the relevance of the input to the individual who processes it.

This disposition to search for relevance is routinely exploited in human communication. Speakers know that listeners will pay attention only to stimuli that are relevant enough and so – in order to attract and hold an audience's attention – they make their communicative stimuli appear at least relevant enough to be worth processing. More precisely, the *Communicative Principle of Relevance* claims that by overtly displaying an intention to inform – producing an utterance or other ostensive stimulus – a communicator creates a presumption that the stimulus is at least relevant enough to be worth processing and moreover, the most relevant one compatible with her own abilities and preferences. This Communicative Principle motivates the following relevance-theoretic comprehension procedure—taken from Wilson and Sperber (2002, p. 13):

> Relevance theoretic comprehension procedure: (a) follow a path of least effort in computing cognitive effects; test interpretive hypotheses (disambiguations, reference resolutions, implicatures, etc.) in order of accessibility; (b) stop when your expectations of relevance are satisfied

The comprehension procedure itself can be seen as a 'fast and frugal' heuristic of the kind mentioned above. In this respect, the relevance theoretic approach diverges from more traditional Gricean accounts of comprehension (see Grice 1989, pp. 30–31), which rationally reconstruct the comprehension process in the form of conscious and reflective inferences about the mental states of others.

Consider again (9) and (10). There are many degrees of anger or disappointment that that speaker might have intended to convey and each of these would be relevant in a different way and yield different implications. Although a neutral tone of voice would cause the hearer least phonological processing effort, it would give him very little guidance on the type of effects he was expected to derive. By contrast, any departure from neutral prosody would increase the hearer's phonological processing effort but would thereby encourage him to look for extra (or different) effects. Which effects should he derive? According to the comprehension procedure above, he should follow a path of least effort, deriving whatever effects are made most accessible in the circumstances by the type of prosodic input used and stopping when he has enough effects to justify the extra effort caused by the departure from the 'expected' prosody.

Another idea often found in the literature is that contrastive stress, like pointing, is a natural highlighting device, used to draw attention to a particular constituent in an utterance. The idea is explored from a relevance-theoretic perspective in Sperber and Wilson (1986/1995, Chap. 4), who build on Grice's work. It follows from the Communicative Principle of Relevance that if two stress

patterns differ in the amounts of processing effort required to interpret them, then the costlier pattern should be used less often and only used in order to create extra, or at least different, effects. Thus, compare the effects on reference assignment of the neutral stress pattern in utterance (11) and the costlier contrastive pattern in (12):

11. Bill insulted Ted and then he *punched* him.
12. Bill insulted Ted and then *he* punched *him*.

A hearer using the relevance-theoretic comprehension procedure in interpreting the second conjunct in (11) should follow a path of least effort in assigning reference, and interpret *he* as referring to Bill and *him* to Ted. This assignment is made easily accessible by syntactic parallelism on the one hand, and encyclopaedic knowledge, on the other. Use of the costlier contrastive pattern in (12) should divert the hearer from this otherwise preferred interpretation towards the alternative, less accessible interpretation on which *he* refers to Ted and *him* to Bill. On this account, contrastive stress is a 'natural' highlighting device, which achieves its effects via the automatic working of the relevance-theoretic comprehension procedure. It does not *encode* anything. Some elements of prosody, however, *do* and it is to those elements I turn in the next section.

7.2.3 What Does Prosody Encode?

As I said earlier, it is a tacit assumption in much of the literature on prosody (and indeed, in the literature on human communication) that prosody conveys the information it does by encoding it. So on the one hand there are three distinct (and probably linguistic) aspects of the prosodic structure of English that contribute to what a speaker means: *tonality* – the chunking of words into groups or phrases; *tonicity* – the location within that phrase of the pitch accent – or *tonic* – a prominent syllable, which typically highlights new information; *tone* – the type of melodic contour on that accent. On the other, there are the natural aspects of prosody which encode information that conveys information about emotional states and attitudes or creates impressions. We have seen that much of what is often treated as governed by either a linguistic or biological code might not be coded at all. Nonetheless, natural codes do exist.

Much early work on prosody and meaning – in what might be called the British School – concerned itself with the meanings of English nuclear tones (*pitch tones* or *melodies*). This is the rising or falling (and rising *and* falling) that occurs on the tonic syllable in an intonation phrase. But assigning meaning to English tones has proved a troublesome business. In general, the more precise one's account, the easier it is to criticise. O'Connor and Arnold (1973) make highly specific claims about meaning. A low fall, according to them, means that a statement is definite and complete insofar as it is a 'separate item of interest'. In addition, it conveys a 'detached, reserved, dispassionate, dull, possibly grim

or surly attitude on the part of the speaker'. But so much depends on the words that such accounts are rendered redundant.

13. I \adore you.

It is now recognised that prosody encodes something much less precise and perhaps hard to pin down in conceptual terms. So rather than a particular tone encoding a concept such as 'detachedness' or 'reservation', the tone encodes information that indicates how the speaker intends the proposition she is expressing to fit in with what she believes the hearer knows or believes at a particular point in the conversation. As Jill House puts it, prosody functions to:

> …guide the listener in how to proceed: how to access the relevant cognitive context within which to interpret the speaker's contribution, how to evaluate that contribution, and how to construct the interaction itself, to enable the communication to take place. (House 2007, p. 369)

House's work (House 1990, 2006) is crucial to the account developed here. She is someone who has constantly tried to bridge the gap between prosody and pragmatics: a phonetician with a real understanding of pragmatic principles; for her, the role of prosody is – at least partly – to *create* the context.

> As communication continues, newly communicated assumptions… or background assumptions which the talk has made accessible, come to the foreground, while others drop into the background. The ever-changing context is thus whatever set of assumptions is active at a given time. Comprehension involves the processing of the new assumptions in the context of the old ones and in the process, the context is updated. (House *ibid.* p. 370)

Hirschberg and Ward (1995, p. 407) propose that the high-rise question contour of English encodes 'that the propositional content of the utterance is to be added to the speaker's and hearer's 'mutual beliefs'… and to question whether the hearer can relate that propositional content to the contents of the hearer's own (unshared) beliefs'. In a recent paper Clark (2007), building on the work of Imai (1998), makes proposals about all the tones of Southern 'Estuary' English: a rise, for example, encodes information to the effect that 'an explicature[2] of the utterance is entertained as an interpretation of a thought of someone other than the speaker at the time of utterance.' Such proposals are vague enough to be worthy of attention, but note that what prosody encodes is often even more vague: affective prosody communicates moods and vague impressions.

We are led to two apparently incompatible claims: on the one hand, the claim that prosodic signals are naturally or linguistically coded; on the other, that they typically create a diffuse impression rather than conveying a determinate message. A *code* is standardly seen as a set of rules or principles pairing signals with determinate messages. How is it possible to maintain both that prosodic signals are coded and that what they convey may be no more than a wide array of weak non-propositional

[2]In relevance theory anything communicated explicitly (as opposed to implicitly – *cf.* Grice's notion of *implicature*) is called an *explicature*.

effects? The answer, I will suggest, is that we need a new notion of coding. In this section of the chapter, I pursue an idea originally proposed by Diane Blakemore (1987, 2002) and applied to different aspects of prosody by Vandepitte (1989), Clark and Lindsey (1990), House (1990, 2006), Escandell-Vidal (1998, 2002) and Imai (1998) and Fretheim (2002). The reason it is hard to pin down what prosody encodes in conceptual terms is that prosody does not encode anything conceptual at all.

The idea is this: if linguistic communication typically involves a combination of decoding and inference, then linguistic signals might be expected to encode information of two distinct types. First, there is regular *conceptual* encoding, where a word (e.g., *dog*) encodes a concept (e.g., DOG), which figures as a constituent of the logical form of sentences in which that word occurs. Second, we might expect to find a form of *procedural* encoding, where a word (or other linguistic expression) encodes information specifically geared to guiding the hearer during the inferential phase of comprehension. The function of such 'procedural' expressions would be to facilitate the identification of the speaker's meaning by narrowing the search space for inferential comprehension, increasing the salience of some hypotheses and eliminating others, thus reducing the overall effort required.

Properly linguistic expressions which have been analysed in procedural terms include discourse connectives, mood indicators and discourse particles (cf. Blakemore 1987, 2002; König 1991; Wilson and Sperber 1993; Hall 2004). So a discourse connective such as *but* encodes a procedure which inhibits a conclusion that might otherwise be drawn; mood indicators – e.g., imperative morphosyntax – encode procedures which facilitate the retrieval of a range of speech-act or propositional-attitude descriptions associated with imperatives; discourse particles such as *please* encode a procedure that facilitates the retrieval of a range of speech-act or propositional-attitude descriptions associated with requests. Properly linguistic prosodic signals (e.g., lexical stress, lexical tone and fully grammaticalised aspects of prosody – perhaps nuclear tones) might be analysed on similar lines, as facilitating the retrieval of certain types of syntactic, semantic or conceptual representation. Thus, the notion of procedural encoding applies straightforwardly to properly linguistic prosodic elements.

Turning to natural signals, there has been some debate about whether interjections such as *oh, ah* and *wow* are properly linguistic. Wharton 2003a surveys the literature and concludes that interjections are best analysed as falling on the natural rather than the properly linguistic side. However, I also argue that interjections are natural signals rather than signs and that they share with discourse connectives and discourse particles the property of encoding procedural rather than conceptual information. On this approach, the function of an interjection such as *wow* might be to facilitate the retrieval of a range of speech-act or propositional-attitude descriptions associated with expressions of surprise or delight, which might be narrowed in context by information derived from prosody, facial expressions, background assumptions, discourse context, etc., and contribute to the speaker's meaning in the regular way, by falling under the relevance-theoretic comprehension procedure.

The line of argument is taken further in Wharton (2003b), who proposes that natural signals such as smiles and other spontaneous facial expressions should also be analysed as encoding procedural rather than conceptual information. The idea can be extended to natural prosody, such as affective tone of voice. On this approach, the function of affective tone of voice – a natural signal – would be to facilitate the retrieval of similar propositional-attitude descriptions to those activated by interjections. This approach makes it possible, on the one hand, to capture the fact that natural signals, interjections and properly linguistic signals, such as mood indicators or discourse particles, all have a coded element and on the other, to explain why what they communicate can sometimes be so nebulous, contextually shaded and hard to pin down in conceptual terms.

If, as Bolinger appears to suggest in the quote on page 4, there is a diachronic dimension to the continuum between display and language, then this continuum may turn out to be a useful tool with which to follow the trail back from arbitrary linguistics expressions to their natural origins: perhaps to prosody too. Chen and Gussenhoven (2003) and Wichmann (2002) suggest that since there is considerable cross-linguistic variation in the way paralinguistic meanings are realised, to a point where they may become heavily stereotyped or even fully grammaticalised, they might also become part of language proper. In Wharton (2009) I suggest ways in which vocalisation might be ranged along such a diachronic (and synchronic) continuum, from an entirely natural gag reflex in which the glottis simply closes, to the related *ugh* [ux] through interjections such as *yugh* [jux] and *yuk* [j∧k] to linguistically productive expressions such as *yucky*, *yuckier* and *yuckiest* (see Padilla Cruz 2009a, b for more discussion of interjections). As I have argued in Wharton (2009), a synchronic version of this continuum is already used in the literature on gesticulation and gesture (McNeill 1992; Kendon 2004).

7.3 Practice

One point on which almost all theorists agree is that it is inappropriate to confer onto prosody the same kind of meaning we generally confer onto words. So rather than a particular tone encoding a concept such as 'politeness', 'surprise', 'disbelief' or 'enthusiasm', the tone encodes information that indicates how the speaker intends what she is saying to fit in with what she believes the hearer knows or believes at a particular point in the conversation. Working from relevance-theoretic perspective, Jill House puts it like this:

> The role of discourse participants may be summarised as follows: the speaker's task is to present his contribution in a way that optimally directs the hearer to the intended interpretation; the hearer's task is to use a combination of linguistic decoding and inferencing to derive hypotheses about the explicit and implicit content of the utterance, and to find the interpretation which seems to be most relevant for the least amount of processing effort. (House *ibid.* p. 371)

Given the support for this kind of claim, it is surprising that the temptation to make such direct claims on the link between prosodic form and meaning – á la

O'Connor and Arnold – is not resisted. But one of the most popular pronunciation course books – Headway Intermediate Pronunciation – makes many such direct links, correlating various intonation patterns with the precise same emotional states listed above. As a former EFL teacher myself, I understand perfectly well that sometimes generalisations simply have to be made. But how useful such generalisations are is open to question. Prosody simply *cannot* be pinned down in conceptual terms and it is potentially misleading to the student to presume it can. Szczepek Reed (2004; Chap. 10) and Couper-Kuhlen (in press) use evidence from research carried out within the framework of interactional linguistics to show how some of even the most basic claims made in the EFL literature about the correlations between pitch change and questions are simply wrong.[3]

Another point of agreement among the theorists is that prosodic meaning is not all cut from the same cloth: Fig. 7.1 above is an attempt to clarify the position. According to the figure, prosody comes in four main varieties. Natural prosody divides into two classes: signs and signals. Non-natural prosody includes those aspects of prosody that are grammatical (phonemes, lexical stress) and also the non-natural aspects of prosody that are cultural rather than linguistic.

A strong version of the claim that a communicative behaviour is 'natural' would entail that it is also universal. But universals in human behaviour are hard to find. It has been claimed that the eyebrow flash is a candidate for a universal communicative behaviour (Eibl-Eibesfeldt 1972) but Paul Ekman (1989, 1992, 1999), himself an advocate of the view that there is a whole range of spontaneous facial expressions that have evolved in humans to reflect a signaller's internal state, denies that the eyebrow flash is used in the United States. To what extent can the fact that some aspects of prosody are natural and hence universal (or almost universal) be exploited by teachers and learners? With caution, probably, but I have had some interesting discussion with both on the degree to which such universals exist. When discussing aspects of this chapter with teachers and learners on a summer course I teach regularly,[4] it was a source of interest to some when I introduced them to Gussenhoven's (2002) *Frequency Code*. This is essentially:

> …Ohala's extension to human speech of Morton's explanation for the widespread similarities in patterns of avian and mammalian vocalisations in face-to-face competitive encounters. Vocalisations by dominant or aggressive individuals are low-pitched, while those by subordinate or submissive individuals are high-pitched. (2002, p. 48)

Some teachers suggested that this might be why when adults try to signal reassurance to children, a relatively high, sometimes rising pitch is used, or even why high pitches in general are associated with positive attitudes. Perhaps an awareness that certain aspects of 'natural' English prosody may be partially (or perhaps wholly) reflected in the natural prosody of the first languages in question might be a strategy worth pursuing.

[3] Those working within the field of interactional linguistics continue to do impressive and insightful work into the role of prosody in sequence- and turn-management (see references in Chap. 10).

[4] The University College London Summer Course in English Phonetics.

Turning to the node on the far right of Fig. 7.1, what can it mean to say that a given behaviour is culturally, rather than linguistically encoded? There are a variety of complications, not the least of which is that there is, of course, a *cultural* element to language itself and the question of what differentiates the cultural from the linguistic is a complex one. But whilst the notion of a cultural, as opposed to linguistic, code may be an unfamiliar one in the analysis of prosody, it is perfectly normal in the study of gesture. Communicators have a whole range of gestures at their disposal and the prosodic continuum from 'natural' to language-specific (Gussenhoven 2002; Pell 2002) is reflected in a gestural continuum suggested by McNeill. At one extreme, there are the entirely natural, non-linguistic gesticulations that are spontaneously used to accompany speech. At the other, there is sign language proper, which is fully linguistic and non-natural in Grice's sense. Between these two extremes lies a range of gestures which, whilst clearly *non-linguistic*, are equally clearly non-natural in Grice's sense. This is the category of culture-specific 'emblems': the British two-fingered insult is one such example, the multi-cultural raised second finger 'salute' another. There may well be elements of prosody which have previously been analysed as 'language-specific' but which are actually cultural stylisations rather than properly linguistic – one example might be the British 'calling contour' – 'hell-*aw*-oo'. Ladd (1996), Gussenhoven (2002, 2004) and Wichmann (2002) suggest there is a considerable cross-linguistic variation in the way what they call 'universal paralinguistic meanings' are realised, to a point where they may become heavily stereotyped or even fully grammaticalised and part of language proper. Among the aims of this chapter has been to draw attention to two possibilities that have perhaps not been so widely considered in the literature: first, not all prosodic inputs are coded at all; and second, the fact that prosodic patterns and their interpretations become stereotyped or vary from language to language is not conclusive evidence that they are *linguistically* coded.

Whilst it is not entirely clear precisely which aspects of prosody are cultural rather than linguistically coded – and nor is it clear to me how we might find out – it seems clear that there *is* a socio-cultural aspect to prosody. Over the past 50 or 60 years, the experience of the 'language learner' has changed immeasurably. At one time, in both method and objective, learning a language was analogous to learning algebra or logic. The method was grammar-translation, the objective grammatical competence. Communicating with the target language being learned, even *speaking* it, was never a consideration. In the modern EFL classroom, the emphasis is very much – to hijack Dell Hymes' (1972) term – about *communicative* rather than just *linguistic* competence.

Some aspects of prosody clearly form part of a sociocultural, rather than the grammatical or structural competence a learner of English is aiming for. Cantarutti (2010) suggests prosody forms part of 'pragmatic and discourse skills, as well as the strategic competence'. It may well be that different aspects of prosody play a role in all the different competences. But if certain aspects of prosody are not even *linguistic*, it could be suggested that the language classroom is not even the place to learn them. Perhaps we should no more be teaching students to mimic some aspects of

prosody than we should be teaching them to mimic other dimensions of English culture: clothing, pastimes or food.

The introductory chapter to a recent book by Morón et al. explains how – building on work by Canale and Swain (1980), Canale (1983) and Bachman (1990) – Celce-Murcia et al. (1995) suggest a more comprehensive model of communicative competence as applied to the second language classroom. It suggests the aim of the modern language learner is to achieve a *range* of competences, of which grammatical or structural competence is only one. These other competences include *sociocultural competence* – the ability to produce utterances that are appropriate to the sociocultural context in which communication takes place (e.g., social factors such as participants' age, gender and power and stylistic factors such as politeness conventions and degrees of formality) – and *strategic competence* – the use of fillers, self-initiated repair or self-rephrasing and appeals for help when the learner does not know a word. This latter competence provides students with:

> […] an ever-present, potentially usable inventory of skills that allows a strategically competent speaker to negotiate messages and resolve problems or to compensate for deficiencies in any of the other underlying competencies. (Celce-Murcia et al. 1995, p. 9)

On the summer course I referred to earlier, I carried out an informal survey in order to obtain a sense of the importance teachers place on such competences. Teachers almost unanimously remarked that grammatical (or structural) competence was the most important to the student. But equal second in the rankings of the various competences was sociocultural competence as defined above. Here is what one teacher had to say:

> I think sociocultural competence cannot be overlooked in language teaching because it leads to appropriate and effective communication. I am a university lecturer in Thailand. My university also offers weekend classes for adult students at undergraduate and postgraduate levels. Some of my students can speak English quite well but many times I've found that they do not use English appropriately.[5]

A problem, of course, is that whilst English speakers certainly have rules or conventions of appropriacy, many learners of English as a second language may be doing so in order to converse with *other* non-native speakers of English, whose rules of appropriacy may differ from the native English speaker's as well as their own. To some extent here we enter the debate articulated in Jenkins (2000), on the degree to which native-like pronunciation is even *desirable* among learners. Jenkins famously suggests that teachers might be better off concentrating on a Core Lingua Franca within which those phonemic distinctions that are unlikely to cause communication problems are abandoned. Whilst the ultimate choice must be left up to the individual student, it seems to me that telling students certain phonetic distinctions are not worth their while attempting is at best unrealistic and at worst patronising.

[5]This quote included with the kind permission of Ms Rachada Pongprairat, Assistant Professor of English, Thepsatri Rajabhat University, Thailand.

In any case, Ostler (2010) suggests that English may be the *last* Lingua Franca in the world and that by 2050 no Lingua Franca will be necessary. ELT learners can breathe a sigh of relief and ELT schools – whose monopoly may soon be coming to an end – had better start planning ahead.

References

Bachman, L. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Barkow, J., L. Cosmides, and J. Tooby, eds. 1995. *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford: Oxford University Press.

Blakemore, D. 1987. *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. 2002. *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge: Cambridge University Press.

Bolinger, D. 1983a. The inherent iconism of intonation. In *Iconicity in syntax*, ed. J. Haiman, 97–109. Amsterdam: John Benjamins.

Bolinger, D. 1983b. Where does intonation belong? *Journal of Semantics* 2: 101–120.

Brazil, D. 1975. *The communicative value of intonation in English*. Birmingham: University of Birmingham.

Canale, M. 1983. From communicative competence to communicative language pedagogy. In *Language and communication*, ed. J. Richards and R. Schmidt, 2–28. London: Longman.

Canale, M., and M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1–47.

Cantarutti, Marina. 2010. Perceiving prosody as a communicative strategy: Prosodic orientation in interaction. In *Proceedings of the XXXV FAAPI Conference: EFL and Art*, Córdoba, Argentina, 23 Sept 2010, Paper 45.

Carston, R. 2002. *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.

Celce-Murcia, M., Z. Dörnyei, and S. Thurrell. 1995. Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics* 6(2): 5–35.

Chen, A., and C. Gussenhoven. 2003. Language-dependence in signalling of attitude in speech. In *Proceedings of Workshop on the Subtle Expressivity of Emotion, CHI 2003 Conference on Human and Computer Interaction*, ed. N. Suzuki and C. Bartneck.

Chevallier, C., I. Noveck, F. Happé, and D. Wilson. 2011. What's in a voice: Prosody as a test for the Theory of Mind account of autism. *Neuropsychologia* 49(3): 507–517.

Clark, B. 2007. 'Blazing a trail': Moving from natural to linguistic meaning in accounting for the tones of English. In *Interpreting utterances; pragmatics and its interfaces. Essays in honour of thorstein fretheim*, ed. R.A. Nilsen, N.A. Appiah Amfo, and K. Borthen, 69–81. Oslo: Novus.

Clark, B., and G. Lindsey. 1990 Intonation, grammar and utterance interpretation. *University College London Working Papers in Linguistics* 2: 32–51.

Couper-Kuhlen, E. in press. Some truths and untruths about final intonation in conversational questions. In *Questions. formal, functional and interactional perspectives,* ed. J.P. de Ruiter. Cambridge: Cambridge University Press.

Cummings, L., ed. 2009. *The pragmatics encylopaedia*. London: Routledge.

Eibl-Eibesfeldt, I. 1972. Similarities and differences between cultures in expressive movements. In *Non-verbal communication,* ed. R. Hinde, 297–312. Cambridge: Cambridge University.

Ekman, P. 1989. The argument and evidence about universals in facial expressions of emotion. In *Handbook of social psychophysiology*, ed. H. Wagner and A. Manstead, 143–164. New York: Wiley.

Ekman, P. 1992. An argument for basic emotion. *Cognition and Emotion* 6(3/4): 169–200.
Ekman, P. 1999. Emotional and conversational nonverbal signals. In *Gesture, speech and sign*, ed. L. Messing and R. Campbell, 45–57. Oxford: Oxford University Press.
Escandell-Vidal, V. 1998. Intonation and procedural encoding: The case of Spanish interrogatives. In *Current issues in relevance theory*, ed. V. Rouchota and A. Jucker, 169–203. Amsterdam: John Benjamins.
Escandell-Vidal, V. 2002. Echo-syntax and metarepresentations. *Lingua* 112: 871–900.
Fretheim, T. 2002. Intonation as a constraint on inferential processing. In *Proceedings of the Speech Prosody 2002 Conference*, ed. B. Bel and I. Marlien, 59–64.
Fridlund, A. 1994. *Human facial expression: An evolutionary view*. San Diego: Academic.
Gigerenzer, G. 2000. *Adaptive thinking: Rationality in the real world*. Oxford: Oxford University Press.
Gigerenzer, G., P. Todd, and The ABC Research Group. 1999. *Simple heuristics that make us smart*. Oxford: Oxford University Press.
Grice, H.P. 1957. Meaning. *Philosophical Review* 66: 377–388.
Grice, H.P. 1967. William James Lectures, delivered at Harvard University, *ms.*
Grice, H.P. 1989. *Studies in the way of words*. Cambridge: Harvard University Press.
Gussenhoven, C. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
Gussenhoven, C. 2002. Intonation and interpretation: Phonetics and phonology. In *Speech prosody 2002: Proceedings of the First International Conference on Speech Prosody*, Aix-en-Provence, ProSig and Universite de Provence Laboratoire Parole et Language, 47–57.
Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
Gussenhoven, C. 2006. Semantics of prosody. In *Encyclopaedia of language and linguistics*, vol. 11, 2nd ed., ed. K. Brown, 170–172. Oxford: Elsevier.
Hall, A. 2004. The meaning of but: A procedural reanalysis. *University College London Working Papers in Linguistics* 16: 199–236.
Halliday, M. 1963. *Explorations in the function of language*. London: Arnold.
Halliday, M. 1967. *Intonation and grammar in British English*. The Hague: Mouton.
Hauser, M. 1996. *The evolution of communication*. Cambridge: MIT Press.
Hirschberg, J., and G. Ward. 1995. The interpretation of the high-rise question contour in English. *Journal of Pragmatics* 24: 407–412.
Hirschfeld, L., and S. Gelman, eds. 1994. *Mapping the mind: Domain specificity in cognition and culture*. Cambridge: Cambridge University Press.
House, J. 1990. Intonation structures and pragmatic interpretation. In *Studies in the pronunciation of English*, ed. S. Ramsaran, 38–57. London: Routledge.
House, J. 2006. Constructing a context with intonation. *Journal of Pragmatics* 38(10): 1542–1558.
House, J. 2007. The role of prosody in constraining context selection: A procedural approach. *In Nouveaux Cahiers de Linguistique Française* 28: 369–383.
Hymes, D. 1972. On communicative competence. In *Sociolinguistics*, ed. J. Pride and J. Holmes. Harmondsworth: Penguin.
Imai, K. 1998. Intonation and relevance. In *Relevance theory: Applications and implications*, ed. R. Carston and S. Uchida, 69–86. Amsterdam: John Benjamins.
Jenkins, J. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.
Kendon, A. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
König, E. 1991. *The meaning of focus particles: A comparative perspective*. London: Routledge.
Ladd, R. 1978. *The structure of intonational meaning*. London: Indiana University Press.
Ladd, R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
McNeill, D. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
Millikan, R. 1984. *Language, thought and other biological categories*. Cambridge: MIT Press.

O'Connor, J., and G. Arnold. 1973. *Intonation of colloquial English*. Harlow: Longman.
Origgi, G., and D. Sperber. 2000. Evolution, communication and the proper function of language. In *Evolution and the human mind: Modularity, language and meta-cognition*, ed. P. Carruthers and A. Chamberlain, 140–169. New York: Cambridge University Press.
Ostler, N. 2010. *The last Lingua franca: English until the return of Babel*. London: Penguin.
Padilla Cruz, M. 2009a. Towards an alternative relevance-theoretic approach to interjections. *International Review of Pragmatics* 1(1): 182–206.
Padilla Cruz, M. 2009b. Might interjections encode concepts? *Lodz Papers in Pragmatics* 5(2): 241–270.
Pell, M. 2002. Surveying emotional prosody in the brain. In *Proceedings of the Speech Prosody 2002 Conference,* ed. B. Bel and I. Marlien, 77–82.
Pierrehumbert, J., and J. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, ed. P. Cohen, J. Morgan, and M. Pollack, 271–311. Cambridge: MIT Press.
Sag, I., and M. Liberman. 1975. The intonational disambiguation of indirect speech acts. Papers from the eleventh regional meeting of the Chicago Linguistics Society. Chicago: Chicago Linguistics Society, 487–497.
Seeley, T. 1989. The honey-bee colony as a superorganism. *American Scientist* 77: 546–553.
Sperber, D. 1996. *Explaining culture: A naturalistic approach*. Oxford: Blackwell.
Sperber, D. 2002. In defense of massive modularity. In *Language, brain and cognitive development: Essays in honor of Jacques Mehler*, ed. E. Dupoux, 47–57. Cambridge: MIT Press.
Sperber, D., and D. Wilson. 1986/1995. *Relevance: Communication and cognition*. Oxford: Blackwell.
van Hooff, J. 1972. A comparative approach to the phylogeny of laughter and smiling. In *Non-verbal communication*, ed. R. Hinde, 209–238. Cambridge: Cambridge University Press.
Vandepitte, S. 1989. A pragmatic function of intonation. *Lingua* 79: 265–297.
Ward, G., and J. Hirschberg. 1988 Intonation and propositional attitude: The pragmatics of L*+H L H%. In *Proceedings of the Fifth Eastern States Conference on Linguistics*, 512–522. Columbus: OSU Department of Linguistics.
Wharton, T. 2003a. Interjections, language and the 'showing'/'saying' continuum. *Pragmatics and Cognition* 11–1(May): 39–91.
Wharton, T. 2003b. Natural pragmatics and natural codes. *Mind and Language* 18(5): 447–477.
Wharton, T. 2009. *Pragmatics and Non-verbal communication*. Cambridge: Cambridge University Press.
Wharton, T. in press. Pragmatics and prosody. In *The Cambridge handbook of pragmatics,* ed. K. Allan and K. Jasczolt. Cambridge: Cambridge University Press.
Wichmann, A. 2002. Attitudinal intonation and the inferential process. In *Proceedings of the Speech Prosody Conference,* ed. B. Bel and I. Marlien, 11–16.
Wichmann, A., and D. Blakemore, eds. 2006. Prosody and pragmatics. Special issue of *Journal of Pragmatics* 38(10): 1537–1541.
Wilson, D., and D. Sperber. 1993. Linguistic form and relevance. *Lingua* 90(1): 1–25.
Wilson, D., and T. Wharton. 2006. Relevance and prosody. *Journal of Pragmatics* 38: 1559–1579.

Chapter 8
Prosody and Feedback in Native and Non-native Speakers of English

Jesús Romero-Trillo and Jessica Newell

8.1 Introduction

The relationship between prosody and pragmatics, or intonation and meaning in everyman's terminology, is relatively clear for speakers of a language although often difficult to explain in practical terms. We all know that the meaning of a sequence of words changes dramatically depending on the subtle features that compose the universe of its prosodic realization. In fact, we can say that the malleability of meaning is a mystifying phenomenon that, when misunderstanding in conversation occurs, puzzles the participants who often swear to have been crystal clear in their intentions and their linguistic realization. The question at stake can be formulated like this: how is it possible that meaning becomes so volatile and unpredictable when it leaves the lips of the speaker? Who or what is to blame when a wrong interpretation of an utterance overrides its apparent meaning transparency?

It can be said that pragmatics deals with the study of meaning in context, although context is always an elusive and subjective composite that tries to shed a holistic, though partial, view on reality. Romero-Trillo (in press a) has identified two main traditions in the study of pragmatics: the first understands pragmatics as a dynamic subject in reality, for instance Mey (1993: 4): 'Pragmatics tells us it's all right to use language in various, unconventional ways, as long as we know, as language users,

J. Romero-Trillo (✉)
Departamento de Filología Inglesa, Facultad de Filosofía,
Universidad Autónoma de Madrid, Madrid, Spain
e-mail: jesus.romero@uam.es

J. Newell
Center for Talented Youth, Johns Hopkins University,
Baltimore, MD, USA

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_8,

what we're doing'. In other words, Pragmatics in this tradition can be the driving force in the transmission of meaning relations in communication. The second tradition describes Pragmatics as a static subject; cf. Leech (1983: 6), 'the study of meaning in relation to speech situations'. In other words, the dynamic tradition paves the way for the novelty of speech whereas the static tries to weave a univocal relationship between language and reality. The first would look for inventive relationships between language, reality and individuals, while the second would try to establish the link between words, meanings and situations.

We believe that the dynamic perspective reflects communication correctly in what can be called 'pragmatic triangulation', defined as 'the process by which the addressor, the addressee and the context constitute a liaison in which the absence or deficient behaviour of any of these three elements would make communication impossible' (Romero-Trillo and Maguire 2011: 226; Romero-Trillo in press b). In this way, pragmatics constitutes a discipline that spans beyond traditional linguistics (phonology, grammar or semantics) but functions in coalition with them. In other words, pragmatics is based on an original linguistic meaning that has suffered some transformation as a result of its contact with reality.

8.2 The Pragmatics of Intonation and the Intonation of Pragmatics

In the process of communication there are different elements that the interlocutors have at their disposal to evaluate and assimilate what is being expressed. These elements have to do with language and kinesics, in conjunction with all the information that the experience of the speakers and hearers have under their belts when they engage in a new conversation. This bulk of information, labelled 'Common Ground', is very important in communication as it establishes the difference between what the mental representation of the speakers understands to be already shared and what they perceive that can be cognitively activated by means of the interaction (Kecskes and Fenghui 2009: 331). From this perspective, interaction is, therefore, is the sum of what is said and what is not said and what is said will necessarily rely on the interplay between the old and new information, also in prosodic terms (Riesco-Bernier and Romero-Trillo 2008).

For the present analysis we will adopt the description of English intonation sketched out by Halliday (1967, 1970) and Cruttenden (1997) *inter alios*. Their approach, the Nuclear Tone Theory, bases its tenets on the study of the tone as a perceptible element for all speakers, as opposed to the Autosegmental Theory (Gussenhoven 1984; Ladd 1996), which analyses the relative comparison of high- or low-pitch accents in an utterance (see Chap. 9, for a more detailed explanation of this theory). The Nuclear Tone Theory considers that there is a hierarchical semantic order in the way meaning is assigned primarily to the tonic element, i.e., the most prominent accented syllable in a tone unit (Halliday 1967, 1970). This element, which has received other denominations such as "focus"

(Brazil 1975) or "pitch accent" (Bolinger 1958), is an element that realizes this prosodic function intertwined with a semantic/pragmatic value, as it carries the most significant meaning in a specific speech segment. This prosodic *cum* pragmatic status has often been underrepresented in many studies of prosody and pragmatics but, in our opinion, opens a new window for the interpretation of language in use that needs further exploration from an acoustic perspective.

We can say that the cooperation between prosody and pragmatics also applies to the other layers of the description of intonation: the tonic is the most prominent element, realizes the nucleus of the information load in the segment and is surrounded by the grammatical elements that scaffold its message. It is important to highlight that we consider the presence or absence of a given element or feature equally important for meaning, as happens in the description of phonemes because one feature in the pragmatic description of prosody is always correlated with another. This is why it is not the same to say:

//Can you/stop it?//

In one go, and with the tonic on 'stop', and:

//Can//you//stop//it?//

With single emphasis and a slight pause after each word. In other words, the system of prosody from a pragmatics perspective is not merely quantitative (description of hertz, milliseconds, decibels, etc…) but also qualitative, with the assignment of pragmatic representations to the quantitative results of prosody. This is precisely the aim of this chapter with regard to the production of feedback elements in native and non-native speakers of English, as we shall see below.

The advantage of this theoretical approach to intonation is that it allows us to describe not only the placement of the tonic (tonicity) but also the other systems that play a role in the description of prosody: tonality (the division of speech into units of meaning – tone groups/units), and tone (the contour direction and shape). Tonality is obviously related to tonicity because the division of information correlates with the acoustic peaks. However, what is important from a methodological perspective is that only one primary tonic can appear in a tone unit. If there is a second, this would always be secondary in pragmatic meaning and acoustic saliency. In fact, when the speaker wants to mark several important tonics English demands the separation of tone groups *via* pauses. It is as if the air pulses were also combined with the cognitive flow of information to create a sort of alignment for the egression of meaning.

The third component, tone, is related to the upward/downward/level movement of the voice pitch in a certain linguistic unit. As mentioned above, this is probably the most salient feature of intonation and is also more clearly related to the basic meanings in communication: questions, statements or indeterminacy. It is interesting to mention how auditory perception reacts positively to these pitch movements, even in the absence of lexical or grammatical elements. In other words, as Romero-Trillo (2001) shows, intonation has the capacity of substituting words by hesitations or interjections in a process that neutralises word classes that we would define as 'grammatical emaciation', e.g.: 'mhm' instead of 'yes' to show agreement. From this perspective,

we can say that prosody is the substance that permeates all grammatical constructions and gives them full-fledged pragmatic meaning in context.

The acoustic studies of the tonic have been essential to understand the relationship between prosody and meaning. For instance, Cooper et al. (1985) and Nooteboom and Kruyt (1987), described the relationship between accentuation of words in context, fundamental frequency and duration. Terken (1991) also discussed the importance of fundamental frequency in the perception of prominence within accented words; and Eefting (1991) underlined the importance of accentuation in the duration of words in context. These studies are very important insofar as they disentangle the knot between the surface form of the message (acoustic cues of tonic words: duration, fundamental frequency and intensity) and the pragmatic force of the message. In fact, although the traditional unit of analysis of the tonic used to be the phoneme or syllable within individual words in some languages like Arabic (De Jong and Zawaydeh 2002), or within clauses or sentences for English or Swedish (Cooper et al. 1985; Eady and Cooper 1986; Heldner and Strangert 2001), we favour the tone unit as the natural environment to study the tonic, due to its pragmatic role in the structure of the message (Romero-Trillo 1994; Riesco-Bernier and Romero-Trillo 2008).

In order to study the relationship between prosody and pragmatic meaning, the acoustic components that we consider relevant are those that can be perceived by the human ear and can be assigned a qualitative interpretation from a communicative perspective. These parameters are the following: the trajectory of the tone (traditionally defined in terms of rising and falling); the frequency (F_0) range (the different frequencies of pitch changes); the duration of the tonic; the intensity (amplitude); and the complexity of the contours (fall-rise, rise-fall, etc.).

8.3 Pragmatic Markers and Prosody

The study of Pragmatic Markers (also called Discourse Markers in some traditions)[1] has traversed different stages in the past years, ranging from a more textual to a more contextual orientation. An advocate of the textual approach is, for example, Fraser (1999: 931) who described Discourse Markers as: 'A class of lexical expressions drawn primarily from the syntactic classes of conjunctions, adverbs and prepositional phrases. In general terms, they signal a relationship between the interpretation of the segment they introduce, S2, and the prior segment, S1. They have a core meaning which is procedural, not conceptual and their more specific interpretation is 'negotiated' in the context, both linguistic and conceptual'. This definition contrasted with Schiffrin's contextual description (1987: 31): 'Markers are sequentially dependent elements that bracket units of talk'. As observed, these definitions depart from divergent positions: while Fraser is more concerned with the

[1] For a detailed explanation of their differences see Romero-Trillo (in press a).

intra-textual relationship of the markers, Schiffrin is primarily interested in their extra/con-textual and interactional relationships.

At present, the term Pragmatic Markers has been widely accepted as the superordinate category for those elements whose meaning resists truth-conditional interpretation (cf. Levinson 1983). In this sense, Discourse Markers are the elements that typically establish the link between a clause and its preceding/forthcoming one from a textual perspective, i.e., they connect adjacent units of information to signal, for example, contrast ('but'), consequence ('so'), or time/logical transition ('then'), etc… For this reason, Discourse Markers need to appear in first position and their primary aim would be to structure information in discourse from a procedural standpoint.

Pragmatic Markers are considered the elements that remain outside the propositional content of a sentence and serve to establish the relationship between the speaker and the message as, for instance, in the case of interjections: 'damn, boy, gosh, jeez…' (Norrick 2009), or in the expression of appraisal or evaluation: 'amazingly, sadly, fortunately…' (Aijmer 2008). As mentioned above, some linguists will opt for the inclusion of Discourse Markers as a category of Pragmatic Markers, while others prefer to distinguish both categories to differentiate procedural meaning from pragmatic meaning.

From a prosodic perspective, Pragmatic Markers are very interesting elements as they appear with multiple intonation realisations and have a distinct prosodic entity when they appear in first position, i.e., they are not pronounced inside the same tone unit as the message they accompany, although they usually appear embedded in the same tone unit if they are at the end (Romero-Trillo in press a). These prosodic features, together with their variation in pragmatic meaning are the reasons why they are difficult for non-native speakers of a language. To illustrate this point, the analysis in the London-Lund corpus of the 54 most frequent Pragmatic Markers carried out by Romero-Trillo (2001) showed, for instance, that the element "well" can appear with the five primary tones described by Halliday (1967), see below for a complete description of the system, although in 62.2% of the cases its function is to show neutral disagreement, while other elements like 'yes' have a multiple realisation (ten different functions). This diversity in the prosodic realisation of a Pragmatic Marker and its multiplicity of functions prompted the definition of Specificity as "the degree of uniqueness that a function shows towards an element, and an element towards a function, in discourse" that can be measured in mathematical terms via the calculation of "Specificity Indexes" of elements and functions (Romero-Trillo 2001: 541).

The question at stake is then: why is there not a univocal relationship between a Pragmatic Marker and its prosodic realisation. In other words, lexical elements always accompany their presence in language with the grammatical constructions that clarify if a certain utterance is a question (auxiliary verb before the subject, 'can you…', 'are you…', or the use of 'do', wh-forms, etc…), or a statement, a command or a conditional. However, the use of Pragmatics Markers presents a great diversity of tone possibilities that makes this matching impossible and the realisation of tone diversity outside any grammatical consideration. We believe

that 'specificity indexes' are the only means to approach the pragmatic meaning of the markers from a statistical perspective, as they show the tendency in their use.

The result of this and the reason why these markers are prototypical elements to analyse intonation in relation to pragmatics, is that they have a preeminent prosodic nature that escapes all grammatical and lexical interpretations. In other words, there is no other class of elements in language that can change their pragmatic meaning with such ease just by modifying their prosodic features. In fact, we can aver that their pragmatic meaning is always assigned by prosody. The question then is: how is it possible to create cognitive categories in our minds on the basis of undefined auditory principles? Our hypothesis in this chapter is that it is possible to describe pragmatic behaviour in acoustic terms and then, assign the functional features that Pragmatic Markers realise in real conversation. As mentioned above, English Pragmatic Markers behave in a certain way in terms of their prosodic features, which is quite complex in comparison with other languages. Therefore, speakers of English as a second or foreign language might be linguistically and – by extension – cognitively handicapped because they may not have access to the cognitive richness that English prosody unfolds with Pragmatic Markers. This cognitive handicap has a clear repercussion, for example, in the management of misunderstanding. In fact, in other studies Romero-Trillo and Lenn (2011) have shown the higher frequency and diversity of the markers used in mixed (English native and non-native) conversations to prevent misunderstanding in comparison with the conversations between native speakers of English.

8.4 Prosody and Feedback in Conversation

Feedback in conversation can be defined as the use of 'linguistic elements to show that [the listener] is following the ideas expressed in the message' (Romero-Trillo 2001: 536). This function was amongst the first to be identified by conversational analysts, although described with different equivalent – and descriptive – terms, for example, 'listener responses' (Dittmann and Llewellyn 1967), in relation to phonemic clauses and its cognitive value (Romero-Trillo 1994), 'accompaniment function', as a support to gaze in interaction (Kendon 1967); or 'back-channelling' (Duncan 1973), as turn-taking elements with an interactional function.

The use of feedback in a conversation is cognitively and interactionally significant, as it shows that the message is processed and accepted by the recipient. The type of feedback also indicates the extent to which the information is perceived as something predictable in the flow of the conversation (neutral feedback) or something completely new or subject to surprise or emotion (emphatic feedback). The difference between neutral and emphatic feedback is realised mainly through prosody and not so much through the selection of the Pragmatic Marker (Romero-Trillo 2001).

In this sense, the use of feedback is essential to maintain the cognitive rhythm of a conversation without disruption. It shows that the meaning flows smoothly from

Table 8.1 Description of the data

File	Duration (min)	No. Feedback PMs
SP-001	15:11	75
SP-002	15:32	89
SP-003	16:18	81
SP-006	15:08	85
SP-014	14:47	85

the speaker to the hearer, so the accompanying vocal signals follow some contextual and prosodic rules according to the principles of Adaptive Management. Adaptive Management can be defined as 'the capacity of a speaker to adapt the grammatical, lexical and pragmatic parameters of discourse through a series of remedial elements and through a principled process, in order to comply with the demands of a new cognitive stage in the conversation via a cognitive standardised process' (Romero-Trillo 2007: 83). In this model, prosody creates context and orients the listener towards a pragmatic liaison with the speaker through the identification of intended *vs*. unintended meaning. For this reason, the speaker-listener relationship will be manifested through the meaning conveyed in the use of feedback by the interlocutors, and through the subsequent pragmatic implications derived from this.

For the study, we have used the acoustic analysis software Praat (Boersma and Weenink 2010) and the intonation contours (primary tones) that will be used to identify the different pitch movements are based on Halliday's model (1967, 1970):

Tone 1: falling
Tone 2: rising
Tone 3: level-rise
Tone 4: rise-fall-rise
Tone 5: fall-rise-fall
Tones 13 and 53: compound tones (1 + 3 and 5 + 3)

The data for the analysis comes from the Spanish section of the *Louvain International Database of Spoken English Interlanguage* (LINDSEI) compiled by Romero-Trillo and Fernández-Agüero. The LINDSEI corpus (Romero-Trillo and María Fernández 2010) is a collection of spoken language data from interviews with intermediate to advanced speakers of English with 11 mother tongues. The interviewers are always native speakers of English. In order to make conversations comparable in terms of topic, all interviews follow the same recording procedure according to the following outline: warming-up activity, informal discussion on a personal experience and a picture description. All the participants in our study are female speakers, of a similar age and cultural background (university students). The Spanish section has a total number of 50 interviews (84,749 words) and for the present study we have selected 5 of these conversations with the following details (Table 8.1):

8.5 Analysis of the Data

The total number of Pragmatic Markers that function as feedback elements is 248 in the case of native speakers and 132 in the case of non-native speakers. An initial auditory perception of the markers in the corpus indicates that there is a difference in the pitch used by the native female speakers (Group A) compared to the non-native female speakers (Group B). In fact, the impression is that the non-native speakers are more assertive in their use of feedback than the native speakers. To verify this impression we analysed the initial and final pitch of the markers in both groups of speakers.

The mean value for the initial pitch of Pragmatic Markers in Feedback by native speakers was 277.88 Hz, and of non-natives 262.80 Hz. The t-test had an almost significant result between both groups ($t = 1.73$; $p = 0.060$), as shown in Graph 8.1.

The mean final pitch in the case of native speakers was 277.98 Hz and for non-natives 249.44 Hz; the results of the t-test showed a significant difference ($t = 3.19$; $p = 0.001$) between both groups (Graph 8.2).

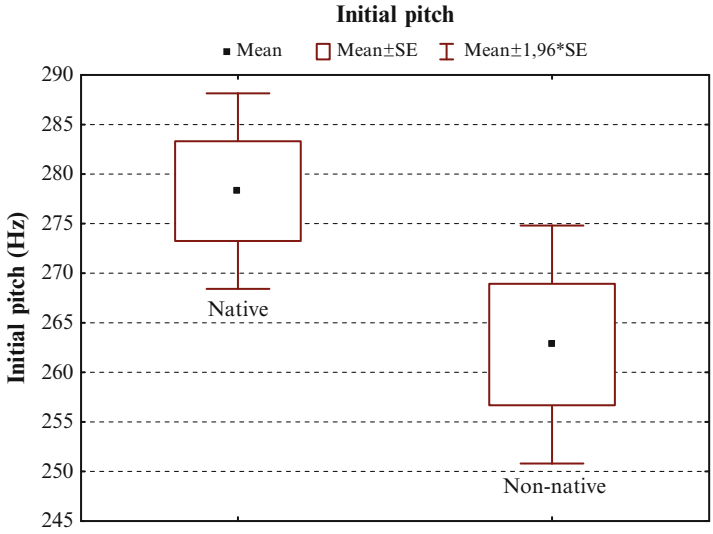
These results show that there is an increasing difference in the pitch level in the use of Pragmatic Markers in the native and non-native group, the final pitch being higher in the case of the native group and therefore, very significant in statistical terms. This is possibly the reason why the native speakers are not perceived as assertive as the non-native speakers.

This difference in the initial and final pitch values in both groups does not imply a blurring of the qualitative value of the tones for non-native speakers. In other words, one could assume that the difference between rising, falling and level tones would be so reduced that there is no margin for differentiation between them in group B. However, this is not the case as the pitch range (the variation between initial and final pitch measures) in both groups is very similar (mean value for A = 65.74 Hz; mean value for B = 66.18 Hz), with no significant statistical differences ($t = 0.17$; $p = 0.95$). This means that all the tone contours (rise, fall, etc.) are acoustically differentiated and proportional in each group and the tones can be audibly differentiated in the non-native speakers.

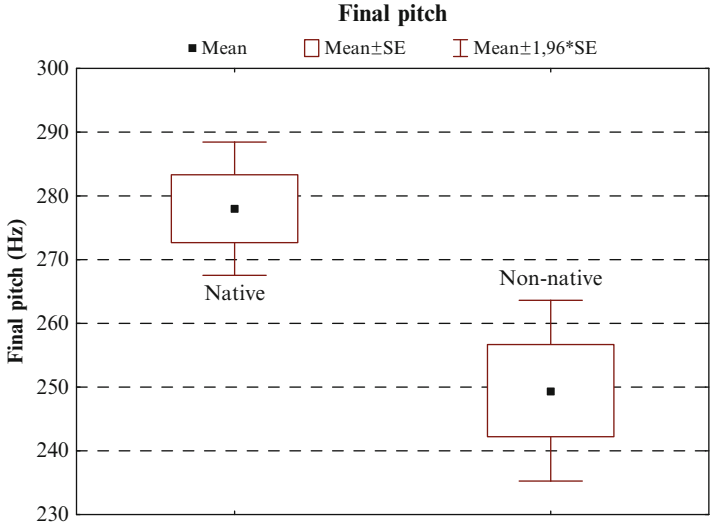
Another auditory perception between both groups is that the Pragmatic Markers are longer in the native than in the non-native speakers. In fact, the mean duration of the markers in the native group is 457.81 ms and in the non-natives 386.16 ms. Again, the t-test showed a significant statistical difference in this measure ($t = 2.78$; $p = 0.005$) (Graph 8.3).

In other words, the overall perception of the use of feedback for both groups shows that native speakers use longer and higher pitched Pragmatic Markers in comparison with non-natives, which indicates that in general terms, feedback may result in being pragmatically unclear, as pitch and duration deviate from the typical native use.

After this description of the general behaviour of the Pragmatic Markers for pitch and duration, we will now proceed to the detailed analysis of the two most frequent elements in both groups: ‘mhm’ (native speakers = 51.61%; non-native

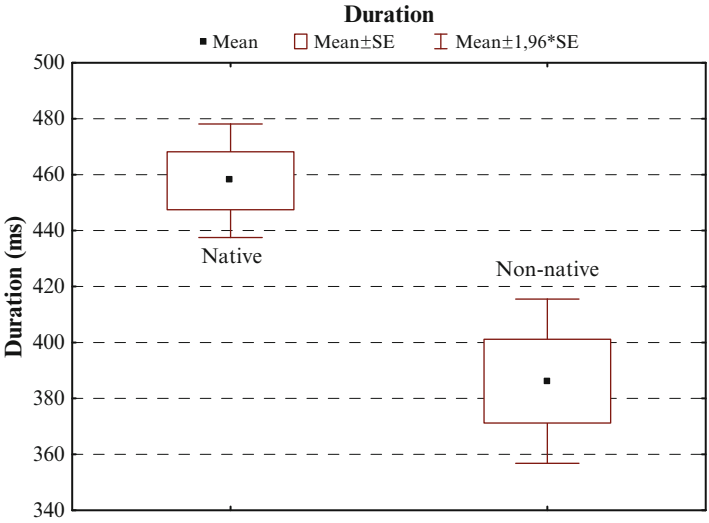


Graph 8.1 Initial pitch of Pragmatic Markers in native and non-native female speakers



Graph 8.2 Final pitch of Pragmatic Markers in native and non-native female speakers

speakers = 15.9%), and 'yeah' (native speakers = 21.7%; non-native speakers = 30.3%). In the description of the element 'mhm' the first noticeable feature is the different mean duration in the native and non-native speakers (Group A = 456.87; Group B = 369.05) with subsequent significant results in the t-test (t = 2.9; p = 0.01).



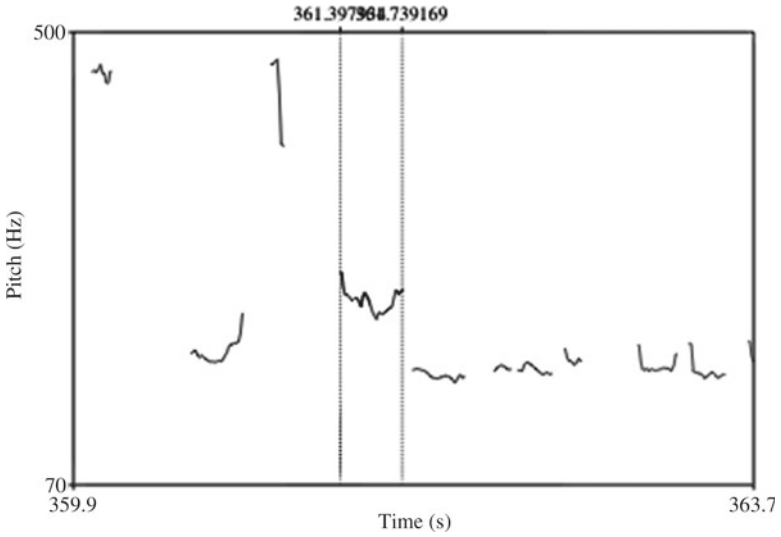
Graph 8.3 Duration of Pragmatic Markers in native and non-native female speakers

As regards the pitch differences, the mean initial pitch value of the marker 'mhm' in the native group is 280.57 Hz and in the non-native group 266.36 Hz. The t-test is statistically significant (t = 2.45; p = 0.01). However, the final pitch mean values for both groups are very similar: Group A = 292.56 Hz; Group B = 266.36 Hz. Nevertheless, in this case the t-test is not statistically significant (t = 1.39; p = 0.16), which shows that the final pitch between both groups is similar. The reason for this unexpected result can be found in the varied realisations of this marker. The native speakers use this element with a complex contour (tone 4: rise-fall-rise, or tone 5: fall-rise-fall) in 63.28% of the cases, while the non-native speakers use these complex tones only in 42.85% of the cases. This means that the difference between both groups also lies in the qualitative perception of the contours because most English speakers use complex contours in which the highest initial pitch level is then supported by a medium pitch level that also functions as a lengthener mechanism. It is interesting that the Spanish speakers' performance of the medium pitch is also similar to the native speakers' (mean value for Group A = 226.69 Hz; Group B = 217.73 Hz), which implies non-significant statistical results (t = 0.38; p = 0.69).

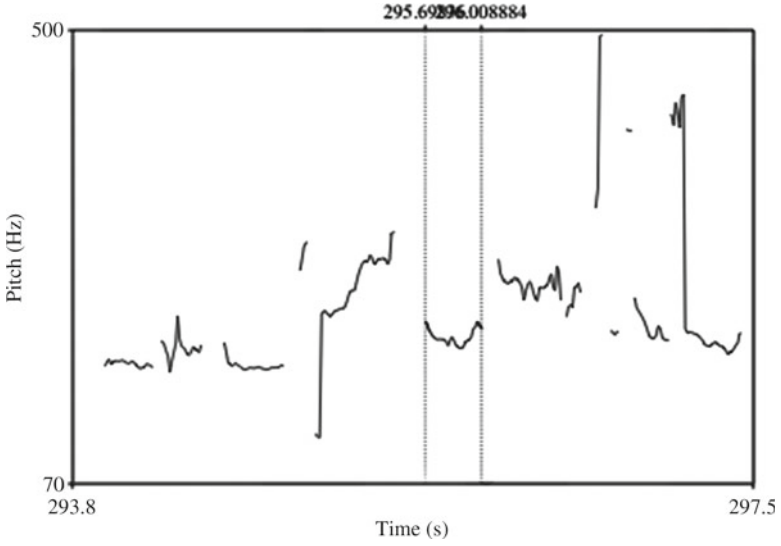
An example of the typical performance of this element can be seen in the following graphs (between vertical lines) (Graphs 8.4 and 8.5):

In these graphs we can observe the longer duration of the English marker, as well as the steeper contrast in the pitch movement in the first case. Both features mark the different acoustic and pragmatic realisations described above in statistical terms.

As regards the Pragmatic Marker 'yeah', the mean duration of the marker in the native speakers is 420.83 ms and in the non-natives 319.89 ms, which renders statistically significant results (t = 3.12; p = 0.002), as is the general case with all the

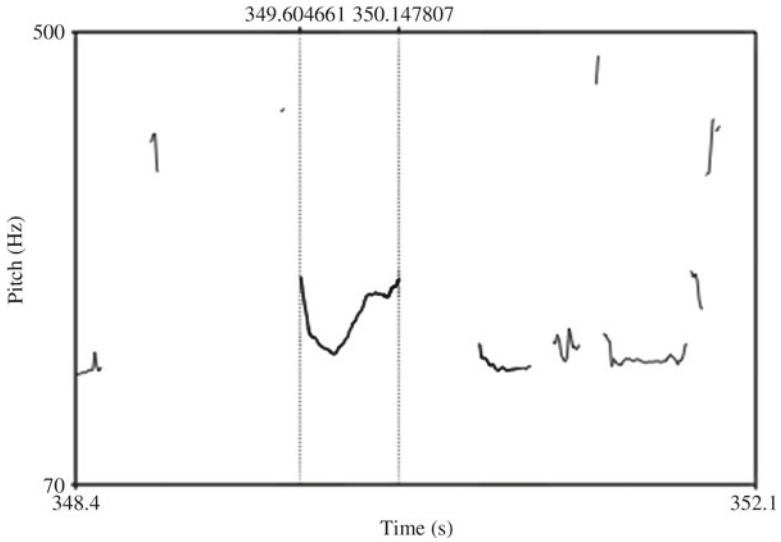


Graph 8.4 Example of 'mhm' uttered by a native speaker

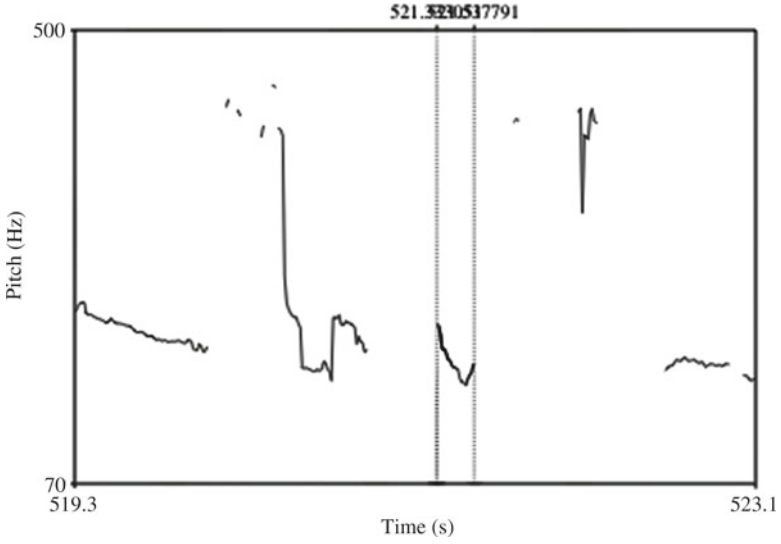


Graph 8.5 Example of 'mhm' uttered by a non-native speaker

markers. In terms of pitch values, the initial pitch has a mean of 280.22 Hz in the native group, which is practically the same as in the non-native group, 288.53 Hz, with no significant results ($t = 42.2$; $p = 0.67$). Again, it is the final pitch that shows a great difference between both groups: 283.52 Hz for the natives and 231.68 for the non-natives ($t = 2.77$; $p = 0.006$).



Graph 8.6 Example of 'yeah' uttered by a native speaker



Graph 8.7 Example of 'yeah' uttered by a non-native speaker

By way of illustration, in the following graphs we present a prototypical example of the marker 'yeah' by a native speaker and by a non-native speaker to see the different realisations (Graphs 8.6 and 8.7).

In sum, the data shows that the element 'yeah' has a longer duration and higher pitch level in final position in the case of native speakers, which indicates a more

tentative and non-assertive meaning than in the use by non-native speakers, who sound more affirmative in the feedback function and whose aim is to show active listenership.

8.6 Conclusions

The present chapter has presented the pragmatic implications of prosodic description in the analysis of Pragmatic Markers that function as feedback. The foundations of the study lie in the awareness that interaction is based not only on the correct transmission of meaning by the speaker, but also on correct processing by the listener. The feedback function precisely serves to verify that the information is being processed and this is why the correct use of the feedback elements is mandatory in oral communication. In fact, when native speakers of English interact with non-native speakers, there are certain features of linguistic performance that may cause misunderstanding or communication deficiencies. In this study we have analysed the frequency and duration behaviour of Pragmatic Markers in conversations between female native and non-native speakers of English, as these two acoustic parameters have implications in the pragmatic interpretations ancillary to prosody. The results of our analysis show a significant difference in the length and pitch of Pragmatic Markers between both groups for all the Pragmatic Markers in the feedback function, then exemplified in the detailed analysis of the two most frequent elements in both groups: 'mhm' and 'yeah'. The result of the analysis shows that the combination of longer markers with the higher final pitch frequency constitutes the reason why these elements behave more interactionally in the native speakers, compared to the somehow more transactional flavour of the markers in the non-natives, in Brown and Yule's terminology (1983).

Pedagogically speaking, we believe that the detailed study of the prosodic features of Pragmatic Markers is fundamental for the correct linguistic behaviour of speakers of English as a foreign/second language, as it can help them not only to sound more native-like in their speech, but also to become more pragmatically accurate in the specific functions that these elements realise. The purpose of this training would be to show that Pragmatic Markers are elements that are outside any grammatical, lexical and syntactic ruling and that they mainly depend on prosody to realise their functional mission in conversation.

References

Aijmer, Karin. 2008. At the interface between grammar and discourse—A corpus based study of some Pragmatic Markers. In *Pragmatics and corpus linguistics: A mutualistic entente*, ed. Jesús Romero-Trillo, 11–36. Berlin: De Gruyter.

Boersma, Paul, and David Weenink. 2010. Praat: Doing phonetics by computer [Computer program]. Version 5.1.07, from http://www.praat.org/.

Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word* 14: 109–149.
Brazil, David. 1975. *Discourse intonation 1/2*. Birmingham: Birmingham University Press.
Brown, Gillian, and George Yule. 1983. *Discourse analysis*. Cambridge: Cambridge University Press.
Cooper, W.E., S.J. Eady, and P.R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America* 77: 2142–2156.
Cruttenden, Alan. 1997. *Intonation*, 2nd ed. Cambridge: Cambridge University Press.
De Jong, K., and B. Zawaydeh. 2002. Comparing stress, lexical focus and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30: 53–75.
Dittmann, A.T., and L.G. Llewellyn. 1967. The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology* 61: 341–349.
Duncan, Starkey. 1973. Towards a grammar for dyadic conversation. *Semiotica* 9: 29–46.
Eady, Stephen J., and William E. Cooper. 1986. Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America* 80: 402–415.
Eefting, Wieke. 1991. The effect of 'information value' and 'accentuation' on the duration of Dutch words, syllables and segments. *The Journal of the Acoustical Society of America* 89: 412–423.
Fraser, Bruce. 1999. What are discourse markers? *Journal of Pragmatics* 31: 931–952.
Gussenhoven, Carlos. 1984. *On the grammar and semantics of sentence accents*. Dordrecht: Foris Publications.
Halliday, Michael A.K. 1967. *Intonation and grammar in British English*. The Hague: Mouton.
Halliday, Michael A.K. 1970. *A course in spoken English*. Oxford: University Press.
Heldner, Mattias, and Eva Strangert. 2001. Temporal effects of focus in Swedish. *Journal of Phonetics* 29: 329–361.
Kecskes, Istvan, and Zhang Fenghui. 2009. Activating, seeking and creating common ground: A socio-cognitive approach. *Pragmatics and Cognition* 17: 331–355.
Kendon, Adam. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26: 22–63.
Ladd, D.Robert. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
Leech, Geoffrey. 1983. *The principles of pragmatics*. London: Longman.
Levinson, Stephen P. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
Mey, Jacob. 1993. *Pragmatics. An introduction*. Oxford: Blackwell.
Nooteboom, Sieb G., and J.G. Kruyt. 1987. Accents, focus distribution and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America* 82: 1512–1524.
Norrick, Neil R. 2009. Interjection as Pragmatic Markers. *Journal of Pragmatics* 41: 866–891.
Riesco-Bernier, Silvia, and Jesús Romero-Trillo. 2008. The acoustics of 'newness' and its pragmatic implications in classroom discourse. *Journal of Pragmatics* 40: 1103–1116.
Romero-Trillo, Jesús. 1994. Ahm, ehm, you call it theme?… A thematic approach to spoken English. *Journal of Pragmatics* 22: 495–509.
Romero-Trillo, Jesús. 2001. A mathematical model for the analysis of variation in discourse. *Journal of Linguistics* 37: 527–550.
Romero-Trillo, Jesús. 2007. Adaptive management in discourse: The case of involvement discourse markers in English and Spanish conversations. *Catalan Journal of Linguistics* 6: 81–94.
Romero-Trillo, Jesús. in press a. Pragmatic Markers. In *Encyclopaedia of applied linguistics*. Oxford: Wiley-Blackwell.
Romero-Trillo, Jesús. in press b. Pragmatic triangulation and misunderstanding: A prosodic perspective. In *Contrastive discourse analysis functional and corpus perspectives,* ed. Maite Taboada, Susana Doval Suárez, and Elsa González Álvarez. London: Equinox.
Romero-Trillo, Jesús, and María Fernández Agüero. 2010. Spanish subcorpus. In *LINDSEI corpus*, ed. G. Gilquin, S. De Cock, and S. Granger. Louvain: Louvain la Neuve Press.

Romero-Trillo, Jesús, and Elizabeth Lenn. 2011. Do you '(mis) understand' what I mean? Pragmatic strategies to avoid cognitive maladjustment. *Journal of English Studies* 9: 235–253.
Romero-Trillo, Jesús, and Laura Maguire. 2011. Adaptive context: The fourth element of meaning. *International Review of Pragmatics* 3: 226–239.
Schiffrin, Deborah. 1987. *Discourse markers*. Cambridge: Cambridge University Press.
Terken, Jacques. 1991. Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America* 89: 1768–1776.

Chapter 9
Early Prosodic Production: Pragmatic and Acoustic Analyses for L2 Language Learners

Heather L. Balog

9.1 Introduction

Prosody refers to the suprasegmental aspects of speech production and includes perceptual features of pitch, length, and loudness. The acoustic correlates to these perceptual features are fundamental frequency (f_o), duration, and amplitude, respectively. Prosodic features overlap segments and serve a variety of purposes, including: contrastive stress, lexical stress, communicative intent, emotion, and attitude (to name a few). The focus of this chapter is primarily on the development of the prosodic system for pragmatic communication and how that development may be influenced by competing language environments. One feature of prosody, intonation, plays a particularly important role in conveying pragmatic function. Intonation relates to the pitch patterns in speech and includes measures such as mean pitch, pitch range, and contour shape.

Pragmatic communication encompasses a variety of factors. For example, communicative intent refers to whether a speaker is making a statement, a command, a request, or asking a question. Pragmatics also includes discourse principles such as topic maintenance, and turn-taking, among others. Within pragmatics, we also have to consider attitude and emotion. Here, attitude relates to how certain the speaker is about what they are communicating and emotion relates to feelings such as happy, sad, or angry.

Development of communication undergoes dramatic and rapid changes during early linguistic stages (e.g., pre-linguistic, first-words, and combinatorial speech). During this period children begin to stabilize their skills in pragmatic function, at the same time as they begin to fine-tune their speech production skills. Specifically, suprasegmental or prosodic speech production development leads to more sophisticated pragmatic communication. As children develop more complex

H.L. Balog (✉)
Balog Speech and Language Center, PLLC, 43824 Joy Rd, Plymouth, MI 48170, USA
e-mail: heather.balog@gmail.com

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_9,

communicative functions, their ability to coordinate ever more complicated prosodic structures, which thereby increases their communicative competence.

It is well-documented that babies produce variable prosodic patterns early in life and that some production patterns may be influenced by the ambient language. How babies definitively master the prosodic characteristics of their ambient linguistic environments remains unclear. We know that mastery of these skills is essential for communication with appropriate grammar and pragmatic function (for a review see Snow and Balog 2002). Even less is known about how babies master the prosodic system when there are competing language environments; that is, how do babies acquire these subtle phonological linguistic cues when they are exposed to bilingual or multilingual (L2) environments. The lack of research on L2 prosodic acquisition opens a door to future research areas that will not only enhance our knowledge of language acquisition processes in general, but also our knowledge of language acquisition when there are competing environments for learning.

The goal of this chapter is to outline: (1) what is known about early pragmatic and prosodic skills in infants and toddlers (referred to as babies throughout the chapter); (2) highlight the research on prosodic acquisition in cross-linguistic and bilingual studies with a special emphasis on intonation; and (3) propose a direction for future research in L2 acquisition based on research with adults and older children. Throughout the chapter, a focus on acoustic analysis of intonation will be emphasized.

To that end, this chapter will define important terminology related to pragmatic communication in the early years and briefly outline normal developmental expectations of these skills. Pragmatic language will be discussed in the context of intonation development with an emphasis on the theoretical and practical considerations for acoustic measurement of intonation produced by babies. Finally, applications of this research to L2 language learners will be proposed.

9.2 Intonation Development in the Context of Pragmatic Communication

At birth, children are born with only the most rudimentary communicative skills. In fact, it is more accurate to argue that newborns are unable to communicate, but rather are limited to signaling. For example, they can indicate basic wants, needs, and emotions with cries, grunts, and gurgles; but these are considered reflexive responses to biological needs rather that truly communicative (e.g., *I am uncomfortable, therefore I cry*). Children are in the prelinguistic stage of communication from birth to approximately 9–12 months. During this time, their gestures and vocalizations develop into more sophisticated forms and begin to shape themselves toward the standards of their ambient language environments. True communication begins only when a child develops intentionality.

Children transition from the prelinguistic to the linguistic stages around 12 months (ranging from 8 to 14 months; Bates 1976). This transition is characterized by a move from the perlocutionary to the illocutionary and locutionary stages.

The transition from preintentional to intentional communication is observed vocally as children begin to produce variegated babbling forms which build the foundation for early meaningful forms (Vihman et al. 1985). Non-verbal characteristics of this transition include the onset of pointing and other communicative gestures, the use of eye gaze to establish joint attention, and other behaviors.

Pragmatic communication is described in a variety of ways for young children and systems used for research purposes range from simple to complex. Simpler categorizations use predetermined situational contexts to categorize pragmatic language (e.g., D'Odorico 1984; D'Odorico and Franco 1991) or broad intention and discourse categories (e.g., statement, command, or request; related or unrelated; Balog and Roberts 2004; Balog et al. 2009). More complex systems offer greater attention to detail for categorizing utterances and a more fine-tuned analysis (Carpenter et al. 1983; Coggins and Carpenter 1981; Coggins et al. 1987; Dore 1975; Ninio et al. 1994; Oller and Eilers 1989; Snow et al. 1996). Regardless of how pragmatic language was measured or judged in the early years, there is ample evidence that intonation contrasts plays a critical role in conveying pragmatic meaning (Snow and Balog 2002).

My own early research explored pragmatic development and uses of intonation in babies using purposefully broad and simple pragmatic categorization systems. Balog and Roberts (2004) studied the development of discourse timing during the second year (i.e., 12–23 months) by examining utterance relatedness as perceived by naïve raters. These raters judged the babies' utterances as either *related* or *unrelated* to the preceding maternal utterance. Utterances for which the raters had 80% or greater agreement were further analyzed for discourse characteristics. Balog and Roberts determined that babies' *related* utterances were characterized by joint attention (as evidenced by eye gaze or other nonverbal signals), topic maintenance (as evidenced by nonverbal or verbal information) and occurred within a 4.25 s timeframe. *Unrelated* utterances, on the other hand, were characterized by topic initiation or a narrowed focus (i.e., a more personal or inward communicative focus), a lack of the joint attention characteristics seen in *related utterances*, and these utterances usually occurred after the 4.25 s timeframe.

Balog et al. (2009) extended the work in Balog and Roberts (2004) by studying prosodic features in slightly more refined but still simple discourse categories. In this study, the *related utterances* were redefined as *co-participatory* (CP) and were described as communicative, close in time to the previous utterance, and to have some of the nonverbal communicative behaviors mentioned above (e.g., joint attention). Balog et al. added *initiation* (IN) as a separate category that was also communicative but not linked to the previous adult utterance like the *co-participatory* utterances. *Initiations* were primarily characterized by a change in topic. The *unrelated* utterance category (Balog and Roberts 2004), was redefined as *narrowed focus* (NF) in Balog et al. and was defined as a non-communicative utterance that carried little or no communicative behavior. These vocalizations seemed to reflect self-talk.

Balog et al. (2009) measured intonation characteristics of babies' productions within these three broad discourse categories (i.e., CP, IN, NF) using accent range (i.e., pitch change on the nuclear tone – discussed below) and contour shape

inventory measures. Both measures were carried out in order to give the most detailed description regarding the babies' intonation abilities. Accent range measures were useful in describing their abilities to changes pitch in their rising and falling contours. The contour inventory measure was useful in measuring the variety of shapes babies were using in different contexts.

Balog et al. (2009) interpreted their data as revealing that babies' learning monolingual English in the second year are capable of using falling contours to differentiate pragmatic categories in their communication but not yet able to use rising contours. Additionally, the babies in that study used a greater variety of contour shapes and types in CP interactions compared to IN and NF interactions, indicating that they may have been more motivated by the ongoing topic of communication and/or have benefited from the adult's support. Further evidence of the benefit of adult support for intonation development was found in Balog (2010). That study revealed that babies produced wider intonation contours immediately following maternal utterances, even though they did not attempt to directly imitate maternal contours.

While Balog and her colleagues demonstrated that babies are limited to some extent in their used of intonation, other studies have shown more flexible developmental patterns (e.g., earlier differentiation of falling and rising contours). The differences between study findings are really dependent on whether prosodic features are being measured in relationship to linguistic meaning or not. Many studies have shown that babies have the ability to produce contrasting intonation within the first year of life. Using descriptors for intonation such as falling, rising or level (e.g., Delack and Fowlow 1978; D'Odorico 1984; Kent and Bauer 1985; Kent and Murray 1982), past research has determined that intonation acquisition reveals influences from the ambient language early on. Kent and Bauer (1985) and Kent and Murray (1982) determined that the predominant contour pattern for 3–12 months old babies learning English was that of falling intonation.

Whalen et al. (1991) demonstrated further evidence of ambient language influence in their cross-linguistic study of five English and five French learning babies. The comparison of English and French is interesting because adult intonation production in those languages is quite distinct. English productions are characterized by predominantly falling contours, while French is characterize by more rising contours. In their study, babies learning English produced 75% of their contours with a falling pattern and French babies' productions were evenly split.

While researchers have successfully demonstrated early language influence, it is clear from some studies that there are early biological constraints that must be addressed. Babies exposed to languages in which rising contours are more prominent (like French) do not match the adult system as easily as those whose ambient language is characterized primarily by falling contours. One explanation for this discrepancy was that while babies were attempting to match the rising patterns of their ambient language, they could not yet override the physiological constraints thought to naturally guide falling productions. This biologically driven preference can be explained by the breath group theory (Lieberman 1967), which proposes that falls are a natural result of reduced subglottal air pressure at the ends of utterance productions. Rising contours require an override of the body's natural tendencies and can only be produced stably once the child reaches a certain point of maturity.

Allen (1983) found further support for a physiological influence on early productions. In his study of six 2-year-olds learning monolingual French, he explained that the persistent use of increased intensity with rising pitch contours was likely driven by the continued reliance on respiratory support for rising contours. He supported this claim by explaining that in adult French productions rising pitch is never produced with rising intensity. More recent work by DePaolis et al. (2008) provided continuing support for biological influences on early prosodic patterns. Their cross-linguistic study of babies learning English, French, Finnish and Welsh showed variability in pitch, intensity and duration measures during the first-word period. They interpreted these findings as demonstrating that at the onset of meaningful speech segmental and suprasegmental characteristics are transitioning toward adult-like stability but that the instability in the system may be due to babies' still immature systems.

Other studies of children younger than 12 months (e.g., D'Odorico 1984; D'Odorico and Franco 1991) have shown that young children are capable of producing differentiated cry patterns by varying their intonation. D'Odorico (1984) used acoustic analyses to measure the intonation characteristics of cry and non-cry vocalizations produced by 4–8 month olds learning Italian ($n=4$) in varying contextual situations. The contexts were meant to establish *a priori* a default pragmatic context and included: (1) mother absent – in which cries were considering *calling* and non-cries were considered *requests*; and (2) mother present – in which cries were labeled *discomfort*. Non-cries were also utterances produced with the mother present but were not analyzed in that study. By measuring mean f_o and contour shape, D'Odorico determined that more rising contours were produced during *request* and *calling* contexts, compared to *discomfort* cries. Additionally, *calls* had greater pitch than *discomfort* cries, which had greater pitch than *requests*. This observation really demonstrated that the children differentiated intonation in varying emotional contexts (Snow and Balog 2002).

In a second study of babies' intonation in the prelinguistic period, D'Odorico and Franco (1991) used acoustic analyses of minimum and maximum pitch to measure productions of five babies from 0 to 11 months of age who were learning Italian. Again, this study utilized communicative situation (i.e., shared experience and independent experience) as *a priori* pragmatic contexts, and made the assumption that intonation production would vary consistently per the situation. Their findings indicated that babies produced utterances characterized by more rises and greater pitch during the shared experience context (this was for 4–9-month-olds only).

Other studies provide foundational glances at prosody and pragmatic development. For example, Marcos (1987) studied 10 children (their ambient language was not clearly specified) ages 14–22 months and reported that intonation differentiated initial requests from repeated requests around 15–16 months. Others have demonstrated developing pragmatic uses of intonation for direct *vs*. assertive communication (Furrow et al. 1990), social and private communication (Furrow 1984), directed or not directed to mother (Galligan 1987).Together these studies indicated that babies do produce variable intonation patterns in their prelinguistic productions but they do not confirm whether there is true pragmatic organization at this stage.

Rather, it is thought that while babies do utilize varying intonation early on, they must reorganize their production patterns around the onset of intentional communication and truly linguistic productions (i.e., the onset of first words). A careful study by Snow (2006) provided evidence for this prelinguistic to linguistic shift in intonation production. In his cross-sectional study of 60 children between 6 and 24 months learning English, Snow demonstrated that intonation development followed a u-shaped (or regressive pattern of development). The data from this research revealed that the intonation patterns produced by the youngest babies (i.e., 6–9 months) were similar to those produced by the older babies (i.e., 18–24 months) and that the children who were around the early first-word stage of development demonstrated significantly less pitch range in their productions. Snow suggested that this regressive pattern was an indication of system reorganization as babies move from pre-intentional to intentional communication.

9.3 Prosody Acquisition for L2 Learners

The consensus across the literature on L2 acquisition is that mastery of the suprasegmental characteristics is integral to native sounding speech production. Moyer (1999) investigated factors influencing the intelligibility of in productions by Mandarin-English L2 learners and stated that while it was unclear how prosodic features impacted intelligibility, it was possible that acquisition of suprasegmentals had a greater impact on intelligibility than accurate segmental production. These assertions supported earlier ones by Pennington and Richards (1986) who wrote that L2 training should focus on all areas of speech production, including suprasegmentals. The importance of suprasegmental acquisition for native-sounding and pragmatically appropriate speech production indicates that bilingual children must not only sort out two distinct segmental systems, but also distinct suprasegmental and pragmatic systems.

There is currently very little research on the prosodic characteristic of young L2 language learners and even less research that studies how these features are impacted by varying pragmatic characteristics within the languages being learned. This issue is further complicated by the variety of languages studied and compared; the ages studied and the methods of analyses used. Table 9.1 presents a summary of papers related to prosody and L2 acquisition. While many more papers were reviewed, this table is limited to those that included children younger than 10 years of age, included a comparison of at least 2 languages and in which there was some acoustic analysis of prosody.

Cross-linguistic studies (e.g., DePaolis et al. 2008; Hallé et al. 1991; Levitt 1993; Vihman et al. 1998; Whalen et al. 1991) have provided an important foundation to our knowledge of prosody acquisition in babies across languages. Together, these

Table 9.1 Selected references and study characteristics for research on child acquisition of prosody as measured by acoustic features

Citation	Study characteristics
DePaulis et al. (2008)	**Study type**: Cross-linguistic **Languages (N)**: American English (10); Finnish (10); French (10); Welsh (10) **Ages**: 10–18 months (coinciding with the onset of first-words) **Pragmatic features**: No **Prosodic features**: analyzed on disyllabic utterances; f_o; intensity; duration **Acoustic analysis**: Yes **Summary**: Each group demonstrated evidence of variability in final syllable lengthening, intensity and stress patterns in their productions; Finnish babies produced more trochaic patterns; Welsh babies produced more iambic patterns
Hallé et al. (1991)	**Study type**: Cross-linguistic **Languages (N)**: French (4); Japanese (4) **Ages**: 18 months **Pragmatic features**: No **Prosodic features**: analyzed on disyllabic utterances; f_o (onset, mean, excursion); duration **Acoustic analysis**: Yes **Summary**: Language groups used distinguishing intonation and duration characteristics by 18 months in words and babbling productions; prosodic features characterized babies' ambient languages
Levitt (1993)	**Study type**: Cross-linguistic **Languages (N)**: French (5); English (5) **Ages**: 7–13 months **Pragmatic features**: No; however, indicated the importance of prosody in communicating pragmatic information **Prosodic features**: f_o; duration; intensity **Acoustic analysis**: Yes **Summary**: Pitch and rhythm influenced by ambient language, but not intensity; suggested a regression for prosody around 9–10 months
Ng et al. (2010)	**Study type**: Bilingual **Languages (N)**: Cantonese/English (86) **Ages**: 5–15 years **Pragmatic features**: No **Prosodic features**: rate of vocal fold vibration; mean speaking f_o; pitch sigma; min and max f_o **Acoustic analysis**: Yes **Summary**: Cantonese characterized by lower mean f_o and reduced pitch range compared to their English productions; indicated that these children used different prosodic patterns in each language

(continued)

Table 9.1 (continued)

Citation	Study characteristics
Trofimovich and Baker (2007)	**Study type**: Bilingual **Languages (N)**: Korean/English (20); English (20) **Ages**: 7–11 years **Pragmatic features**: No **Prosodic features**: stress timing; peak alignment; speech rate pause frequency; pause duration **Acoustic analysis**: Yes **Summary**: Children with less than 1 year L2 experience had not yet acquired native suprasegmental skill; children with 11 or more years L2 experience had acquired all native suprasegmental skills except speech rate
Whalen et al. (1991)	**Study type**: Cross-linguistic **Languages (N)**: French (5); English (5) **Ages**: 5–13 months **Pragmatic features**: No **Prosodic features**: Intonation contour (perceptual); f_o **Acoustic analysis**: Yes **Summary**: French babies used more rising intonation contours compared to English babies; English babies used more falling contours compared to French babies

papers defined prosodic features in 5–13 months old babies acquiring American English, French, Finnish, Welsh and Japanese. The consensus across these studies was that early prosodic patterns in babies' productions indicated an early influence of babies' ambient language environment. Many of the studies also mentioned that productions were still characterized by a great deal of variability in productions, especially rising contours. This variability was typically explained as a natural developmental process that was likely influences by the physiological constraints of an immature system (discussed earlier) and/or a regression associated with the onset of other language skills (Levitt 1993).

When viewed in conjunction with monolingual studies of prosodic acquisition (reviewed earlier), we begin to see that while prosodic features are produced very early in life, they undergo a developmental process similar to that of segmental features over the course of the first few years of life. That is, there is not a universal linguistic use of prosody, rather a universal reflexive use of prosodic features that slowly acclimate to the ambient language environment as babies move from preintentional to intentional communication.

Two studies reviewed compared prosodic characteristics in older bilingual children (Ng et al. 2010; Trofimovich and Baker 2007). Trofimovich and Baker compared 20 bilingual Korean/English learning children with varying degrees of L2 experience (ages 7–11 years old) with 20 monolingual English learning children. In that study, a delayed sentence imitation task was used to elicit the speech samples. This task resulted in six declarative sentence productions from each child in response to an experimenter's question. They used this procedural design to help ensure

similar (even identical) responses from the children and to minimize the children's difficulty with lexical, syntactic and pragmatic language features. Acoustical analyses were completed on the productions and included features known to be difficult for L2 acquisition of English and to play a critical role in fluency. The features included measures of rhythm (relative to stress timing), measures of pitch (specifically the highest f_o value on the accented syllable), speech rate and measures of pause frequency and duration. The children in their study with only 1 year of L2 English experience had not yet acquired native suprasegmental skills, whereas their subjects with 11 or more years of L2 English experience did demonstrate native like prosodic features (with the exception of speech rate, which remained slower). These findings were interpreted as an indication that like segmental features, acquisition of L2 suprasegmental features is a process that requires time.

Another bilingual study observed prosodic features in 86 Cantonese-English L2 learners (ages 5–15 years old; Ng et al. 2010). The researchers analyzed 2-minute spontaneous speech samples in each language. Acoustic features of f_o (including mean f_o, min f_o, and max f_o) were analyzed using Praat (Boersma and Weenink 2010). Overall pitch declined with age (i.e., that older children produced lower pitch than the younger children). More interestingly, however, was their finding that the children in their study used a significantly lower pitch when speaking Cantonese. This finding was interpreted as evidence of language differentiation using prosodic features.

Interestingly, none of the studies discussed above explored pragmatic features of language development in conjunction with prosodic features; although Levitt (1993) did mention that use of appropriate prosody was an important factor in effectively communicating pragmatic language. Speer and Ito (2009) emphasized the importance of more cross-linguistic work in order to better understand the role of prosody acquisition and its relationship to other language features (such as syntax, semantics and pragmatics). They also stated that investigating early pragmatic function is particularly tricky, especially in very young children because they lack the cognitive sophistication required to efficiently manage all the linguistic systems simultaneously.

9.4 Acoustic Analysis Techniques

Acoustic analysis of babies' productions is undoubtedly one of the most interesting ways in which to explore acquisition of prosody, whether in monolingual or bilingual learners. The very fact that acoustic analysis provides the researcher a measurable way to define differences makes it particularly appealing to the exploration of language differences across and within individuals and languages. In general, studies focus on the perceptual features of pitch, length and loudness, which are realized acoustically as fundamental frequency (f_o), duration and intensity (amplitude). The features most closely associated with pragmatic language and intonation are pitch and duration. In terms of pitch, intonation can be defined categorically by contour shape, f_o range, mean f_o, minimum and maximum f_o, to name a few (e.g., Balog and Snow 2007; Balog et al. 2009; Cruttenden 1997; D'Odorico 1984; D'Odorico and

Franco 1991; Hallé et al. 1991). Duration measures related to intonation are likely to include measures of speaking rate, final syllable lengthening, pause duration and possibly, rate of pitch change.

Before acoustic analysis is undertaken, the theoretical framework that guides a researcher's approach to prosody should be carefully considered because that framework will guide important decisions for how acoustic analyses are implemented. This is especially true for analyses of baby productions and when attempting to link intonation and pragmatic language skills. It is not clear in most of the studies reviewed in Table 9.1 that a particular theoretical framework for intonation was used. However, there are hints in the procedural descriptions that indicate these researchers were thinking carefully about how to measure intonation. For example, DePaolis et al. (2008) and Hallé et al. (1991) specified that they measured intonation on disyllabic *versus* mono-syllabic utterances.

Where intonation measures are made has an impact on what we learn and likely depends greatly on what is known about the pragmatic and grammatical features of the languages being compared. Snow and Balog (2002) reviewed two competing theories of intonation in the context of pragmatic communication. Their research over time has focused on the acquisition of intonation features in babies up to two years old and has been driven partly by an interest in defining the relationship between intonation and early communication intentions (one aspect of pragmatic language). Their review compared the Tone and Break Indices system (ToBI; Beckman and Ayers 1994; Pierrehumbert 1980; Pierrehumbert and Hirschberg 1990), which comes from autosegmental theory of intonation (Gussenhoven 1984; Ladd 1996) and the nuclear tone theory (Cruttenden 1997) in terms of their applications to developmental prosody.

Autosegmental approaches to intonation seem to be the most currently used, at least in intonation research with English. To summarize briefly, autosegmental approaches view intonation as a sequence of high and low pitch accents (marked as H and L), phrasal accents (marked as -) and edge tones (marked as%). Within ToBI (Beckman and Ayers 1994; Pierrehumbert 1980; Pierrehumbert and Hirschberg 1990), the tonal tier represents the sequence of pitch changes across the utterance. The break index tier marks word boundaries within the middle of the phrase (1), accented words not at an edge boundary (2), phrasal tones (3) and boundary tones (4). Direction of the intonation contour is thought to carry discourse meaning such as *background* (Gussenhoven 1984). For example, direction is used to differentiate new (marked as H) and old information.

The nuclear tone theory (a more traditional approach to the analysis of intonation) focuses more specifically on the nuclear accent of an utterance (Cruttenden 1997). Utterances consist in one or more intonation groups, which are defined by clausal boundaries. Each intonation group has a nuclear accent, which is the primary stress syllable (Crystal 1986; Vanderslice and Ladefoged 1972) within the tone group and is generally produced toward the right edge of that tone group. The nuclear tone begins at the nuclear accent or the syllable in which there is primary stress and continues to the final boundary of the utterance.

Cruttenden (1997) eloquently highlights the tonal factors that can be measured from the nuclear tone and described their purposes. These tonal factors are

direction, accent range and complexity of the contour. *Direction* refers to whether the primary direction of the contour is falling or rising. *Accent range* refers specifically to the amount of pitch change that occurs during production of the nuclear tone. Accent range is calculated by determining the minimum and maximum pitch points for the contour and is best described using a logarithmic scale, such as cents or semitones (Burns and Ward 1982). Finally, contour *complexity* refers to whether there is a change in direction during the production of the contours (e.g., rise-*fall*-rise).

The nuclear tone is the place in the utterance where the most salient grammatical (e.g., boundary) and pragmatic (e.g., attitude, intention, etc.) information is carried. Boundaries are strongly cued by phrase final lengthening of the nuclear tone (Snow and Balog 2002). Contour direction of the nuclear tone is very important within this approach as it is in this part of an utterance that English speakers convey intentionality and attitude. Intentionality is differentiated with falling and rising contours; falls used for statements, commands, requests, *wh*-questions and rises used for *yes/no* questions. More broadly, rises are thought to convey the attitude of less certainty or more openness and falls are thought to convey more certainty or closed attitudes (Cruttenden 1981).

Regardless of the theoretical approach preferred by individual researchers, several factors have to be carefully considered before intonation can be analyzed and implications for its development and relations to pragmatic language can be made. First and foremost, researchers must define in what part of the utterances they are measuring intonation. Historically, studies have intonation acquisition have failed to definitively describe where within an utterance intonation was measured within an utterance (Snow and Balog 2002). At minimum, researchers must define whether intonation was measured across the whole utterance (multi-syllables) or on a smaller portion of the utterance (e.g., the nuclear tone). Then, they must use the proper terminology for these measures. For example, *register* defines the overall pitch height of an utterance. *Key* defines the pitch change that occurs over an entire utterance; whereas *accent range* (described above) defines pitch change over a smaller, more specific portion of the utterance. How and where intonation is measured may vary depending on the languages being compared.

Secondly, researchers must define the relationship between prosodic features and pragmatic language use for the languages being studied. Pragmatic meaning as expressed through intonation was discussed in the light of English language acquisition above but the way in which intonation and pragmatic communication intermix differs across languages and must be taken into consideration when studying L2 acquisition of prosody.

9.5 Implications for Future Research

Clearly, there is tremendous opportunity for future research in the area of intonation and pragmatic language acquisition in L2 language learners. Currently, there is simply not enough research to make definitive claims. Future work in this area must be carefully planned and carried out. Intonation measures must be thoughtfully

conducted and defined in order to maximize their contribution to the current knowledge base. To that end, researchers must carefully define how they measured both prosodic and pragmatic features. Future work should clarify and define the theoretical frameworks from which to build procedures. For baby productions, it is my opinion that the simpler the framework the better. Theories that rely on meaningfulness in the production will not adequately help us measure acquisition during the earliest years of production. To that end, the nuclear tone approach has been proven to be useful with baby utterances, at least for English (see Balog 2010; Snow 2006, among others).

Secondly, careful study of the adult patterns for pragmatic communication must be defined. This will differ across languages and therefore makes it of primary importance to carefully outline the pragmatic similarities and differences between the specific languages being compared. Each combination of languages being acquired likely has a unique influence over how these two systems are acquired.

Ideally, future studies will be conducted longitudinally within a play format to give us an idea of the most naturalistic use of prosodic and pragmatic features in young L2 learners. This is no small undertaking, as it is difficult to recruit families with small children for a long-term commitment to research. Therefore, the research community will have to be open-minded and willing to accept studies with small sample sizes and imperfect data. It is also recommended that future studies be conducted over a protracted period of time, beginning with the pre-linguistic period and following babies into fully grammatical productions. Studies of children between 6 months and 5 years would help solidify L2 acquisition in children with L2 exposure from a very early age and would enable comparison to children who acquire a second language at a later point in development.

This is time-consuming work but the technology available for research in these areas is much improved in recent years. For example, Praat (Boersma and Weenink 2010) is considered to be a state-of-the-art acoustic analysis system and is available without charge. Secondly, digital recorders and improved microphone quality make recording and downloading to the computer for acoustic analysis simple and fast. Using these new advances in technology, the measurements that will most likely add to our knowledge base include accent range, contour direction and shape descriptions and durational measures (e.g., phrase final lengthening). Rather than reinventing the wheel entirely, research such as that described above in Table 9.1 and others can be used as a compass toward creating a foundation for this work in the future.

References

Allen, G.D. 1983. Some suprasegmental contours in French two-year-old children's speech. *Phonetica* 40: 269–292.

Balog, H.L. 2010. A comparison of maternal and child intonation: Does adult input support child production? *Infant Behavior & Development* 33: 337–345.

Balog, H.L., and F.D. Roberts. 2004. Perception of utterance relatedness during the first-word period. *Journal of Child Language* 31: 837–854.

Balog, H.L., and D. Snow. 2007. The adaptation and application of relational and independent analyses for intonation production in young children. *Journal of Phonetics* 35: 118–133.
Balog, H.L., F.D. Roberts, and D. Snow. 2009. Discourse and intonation development in the first-word period. *Enfance* 3: 293–304.
Bates, E. 1976. *Language and context: The development of pragmatics*. New York: Academic.
Beckman, M. E., and G.M. Ayers. 1994. Guideline for ToBI labeling. Version 2. Ohio State University, Columbus.
Boersma, P., and D. Weenink. 2010. Praat: Doing phonetics by computer [Computer program]. Version 5.1.43. Retrieved 4 Aug 2010 from http://www.praat.org/.
Burns, E.M., and W.D. Ward. 1982. Intervals, scales, and tuning. In *The psychology of music*, ed. D. Deutsch, 241–269. New York: Academic.
Carpenter, R.L., A.M. Mastergeorge, and T.E. Coggins. 1983. The acquisition of communicative intentions in infants eight to fifteen months of age. *Language and Speech* 26: 101–116.
Coggins, T.E., and R.L. Carpenter. 1981. The communicative inventory: A system for observing and coding children's early intentional communication. *Applied PsychoLinguistics* 2: 235–251.
Coggins, T.E., L.B. Olswang, and J. Guthrie. 1987. Assessing communicative intents in young children: Low structured observation or elicitation tasks? *The Journal of Speech and Hearing Disorders* 52: 44–49.
Cruttenden, A. 1981. Falls and rises: Meanings and universals. *Journal of Linguistics* 17: 77–91.
Cruttenden, A. 1997. *Intonation*, 2nd ed. Cambridge: Cambridge University Press.
Crystal, D. 1986. Prosodic development. In *Studies in first language development*, ed. P.J. Fletcher and M. Garman, 174–197. New York: Cambridge University Press.
D'Odorico, L. 1984. Non-segmental features of prelinguistic communications: An analysis of some types of infant cry and non-cry vocalizations. *Journal of Child Language* 11: 17–27.
D'Odorico, L., and F. Franco. 1991. Selective production of vocalization types in different communication contexts. *Journal of Child Language* 18: 475–499.
Delack, J.B., and P.J. Fowlow. 1978. The ontogenesis of differential vocalization: Development of prosodic contrastivity during the first year of life. In *The development of communication*, ed. N. Waterson and C. Snow, 93–110. New York: Wiley & Sons.
DePaolis, R.A., M.M. Vihman, and S. Kunnari. 2008. Prosody in production at the onset of word use: A cross-linguistic study. *Journal of Phonetics* 36: 406–422.
Dore, J. 1975. Holophrases, speech acts and language universals. *Journal of Child Language* 2: 21–40.
Furrow, D. 1984. Young children's use of prosody. *Journal of Child Language* 11: 203–213.
Furrow, D., W. Podrouzek, and C. Moore. 1990. The acoustical analysis of children's use of prosody in assertive and directive contexts. *First Language* 10: 37–49.
Galligan, R. 1987. Intonation with single words: Purposive and grammatical use. *Journal of Child Language* 14: 1–21.
Gussenhoven, C. 1984. *On the grammar and semantics of sentence accents*. Dordrecht, Holland: Foris Publications.
Hallé, P.A., B. de Boysson-Bardies, and M.M. Vihman. 1991. Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech* 34: 299–318.
Kent, R.D., and H.R. Bauer. 1985. Vocalizations of one-year-olds. *Journal of Child Language* 12: 491–526.
Kent, R.D., and A.D. Murray. 1982. Acoustic features of infant vocalic utterances at 3, 6 and 9 months. *Journal of the Acoustical Society of America* 72: 353–365.
Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
Levitt, A.G. 1993. The acquisition of prosody: Evidence from French- and English-learning infants. In *Developmental neurocognition: Speech and face processing in the first year of life*, ed. B. de Boysson-Bardies, S. De Schonen, P. Jusczyk, P. McNeilage, and J. Morton, 385–398. Dordrecht: Kluwer Academic Publishers.
Lieberman, P. 1967. *Intonation, perception and language*. Cambridge: MIT Press.
Marcos, H. 1987. Communicative functions of pitch range and pitch direction in infants. *Journal of Child Language* 14: 255–268.

Moyer, A. 1999. Ultimate attainment in L2 phonology: The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21: 81–108.

Ng, M.L., G. Hsueh, and C.S. Leung. 2010. Voice pitch characteristics of Cantonese and English produced by Cantonese-English bilingual children. *International Journal of Speech-Language Pathology* 12: 230–236.

Ninio, A., C.E. Snow, B.A. Pan, and P.R. Rollins. 1994. Classifying communicative acts in children's interactions. *Journal of Communication Disorders* 27: 157–187.

Oller, D.K., and R.E. Eilers. 1989. A natural logic of speech and speech-like acts with developmental implications. *First Language* 9: 225–244.

Pennington, M., and J. Richards. 1986. Pronunciation revisited. *TESOL Quarterly* 20: 207–225.

Pierrehumbert, J. 1980. The phonology and phonetics of English intonation. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Pierrehumbert, J., and J. Hirschberg. 1990. The meaning of intonational contours in the interpretation of discourse. In *Intentions in communication*, ed. P.R. Cohen, J. Morgan, and M.E. Pollack, 217–311. Cambridge: MIT Press.

Snow, D. 2006. Regression and reorganization of intonation between 6 and 23 months. *Child Development* 77: 281–296.

Snow, D., and H.L. Balog. 2002. Do children produce the melody before the words? A review of developmental intonation research. *Lingua* 112: 1025–1058.

Snow, C., B.A. Pan, A. Imbens-Bailey, and J. Herman. 1996. Learning how to say what one means: A longitudinal study of children's speech act use. *Social Development* 5: 56–84.

Speer, S.R., and K. Ito. 2009. Prosody in first language acquisition – Acquiring intonation as a tool to organize information in conversation. *Language and Linguistics Compass* 3: 90–110.

Trofimovich, P., and W. Baker. 2007. Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied PsychoLinguistics* 28: 251–276.

Vanderslice, R., and P. Ladefoged. 1972. Binary suprasegmental features and transformational word-accentuation rule. *Language* 48: 819–838.

Vihman, M.M., M.A. Macken, R. Miller, H. Simmons, and J. Miller. 1985. From babbling to speech: A re-assessment of the continuity issue. *Language* 61: 397–445.

Vihman, M.M., R.A. DePaolis, and B.L. Davis. 1998. Is there a "trochaic bias" in early word learning? Evidence from infant production in English and French. *Child Development* 69: 935–949.

Whalen, D.H., A.G. Levitt, and Q. Wang. 1991. Intonation differences between the reduplicative babbling of French- and English-learning infants. *Journals of Child Language* 18: 501–516.

Chapter 10
Prosody in Conversation: Implications for Teaching English Pronunciation

Beatrice Szczepek Reed

10.1 Introduction

This chapter explores some of the implications of conversation analytic research into prosody-in-interaction for the teaching of English pronunciation. In doing so, it contributes to recent efforts to relate Conversation Analysis (CA) to other research fields, in this case English language teaching (Cf. Schegloff et al. 2002; Wong 2002; Richards and Seedhouse 2005; Seedhouse 2005; Bowles and Seedhouse 2007; Wong and Waring 2010; Hall et al. 2011). However, the overall aim of this chapter is not to apply individual research findings to the design of specific exercises, even though several of the findings below can, and should, be applied in a concrete manner in the future. Rather, it is proposed here that conversation analytic findings concerning prosodic phenomena suggest a perspective on prosody and perhaps on language as such, that is less led by a focus on 'discourse functions' of individual linguistic forms (prosodic or otherwise), and more strongly influenced by an emphasis on language as a set of practices for interactional negotiation, and conversation as a collaborative social accomplishment.

CA has provided a perspective on the phonetics and prosody of naturally occurring talk for at least the past 30 years. Following its emergence in the 1960s, when CA was squarely situated in the field of sociology,[1] linguists too have become aware of its potential for the study of everyday language use. Subsequently, the linguistic branch of CA emerged as *Interactional Linguistics* (Selting and Couper-Kuhlen 2001). In this field, two main areas of study have so far proved particularly

[1] Cf. Sacks (1992). For comprehensive overviews of CA see, for example, Psathas (1995), Liddicoat (2007), ten Have (2007), Hutchby and Wooffitt (2008).

B. Szczepek Reed (✉)
Department of Education, University of York, Heslington, York, UK
e-mail: beatrice.szczepek.reed@york.ac.uk

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_10,

fruitful: grammar/syntax (Cf. Ford 1993; Ochs et al. 1996; Bybee and Noonan 2001; Couper-Kuhlen and Kortmann 2002; Golato 2005; Betz 2008), and phonetics/phonology (Couper-Kuhlen 1986; Kelly and Local 1989; Selting 1995; Couper-Kuhlen and Selting 1996; Couper-Kuhlen and Ford 2004; Barth-Weingarten et al. 2009; Ogden 2009; Barth-Weingarten et al. 2010; Szczepek Reed 2010a). In both areas, the conversation analytic approach has brought with it a new way of studying linguistic events. Many previous notions have been deconstructed and/or re-defined. More importantly, an attempt is being made to study language from the perspective of those who use it, i.e., the conversational participants themselves and with as little recourse as possible to pre-existing linguistic theory. In particular, the common linguistic distinction between 'underlying' and 'surface' patterns has been abolished. Instead, what is being analysed in CA/interactional linguistic research is the empirically observable, naturally occurring behaviour of interactants. Analytic interpretations regarding language-in-interaction are being based exclusively on evidence of their reality for participants.

In the field of interactional linguistic enquiry into phonetics/phonology, two main foci have dominated research to date: the role of prosody for turn- and sequence management (Cf. Local et al. 1985, 1986; French and Local 1986; Local and Kelly 1986; Couper-Kuhlen 1991, 2004; Local 1992; Selting 1996a; Wells and Peppè 1996; Schegloff 1998; Wells and Macfarlane 1998; Fox 2001; Szczepek Reed 2004); and its role in the accomplishment of specific conversational actions, such as news and news receipts (Freese and Maynard 1998; Local 1996); quoting and reported speech (Couper-Kuhlen 1996; Klewitz and Couper-Kuhlen 1999); repair initiation (Selting 1996b) and turn increments (Walker 2004), to name only a few. Findings from this research have so far not been related to English pronunciation teaching, even though a number of opportunities for conceptual adjustment present themselves as a result of them. Two of these opportunities are presented in some detail in the sections below.

The following section explores the relationship between prosodic forms on the one hand, and grammar and discourse functions on the other. Recent research shows that such a relation is necessarily complex. As a result, teaching direct connections between specific intonation patterns and interactional 'meanings' is oversimplified at best, if not plainly misleading. The subsequent section pursues a similar goal in deconstructing the notion that specific prosodic cues have any function outside the local interactional context. The section shows that in natural conversation, a fundamentally relevant issue for participants is not whether they are using a particular prosodic format in and of itself, but whether they are doing prosodically the same as, or something different from what an immediately previous speaker has been doing prosodically. The final section outlines implications of these findings for the teaching of English pronunciation and remarks on the interrelation between research findings and their applicability to the language classroom.

10.2 Prosody and Turn Taking: The Form-Function Relationship

Much of the early research on prosody in conversation had a strong interest in turn taking and the role that prosodic features play in its achievement. Some of the most influential studies of prosody in interaction were conducted against the background of the British school of intonation,[2] which involves a strong form-function relationship between, particularly, intonation patterns and discourse meaning.[3] However, while in the early days of the research field such a connection was perhaps sometimes sought, it soon became clear that detailed observation of conversational data does not warrant a straightforward connection between prosodic events and meanings or even functions.

A first realization that emerged was that conversational actions are implemented by participants' orientation to many simultaneously employed interactional practices, some of which are prosodic, others syntactic, lexical, pragmatic, gestural, etc. Even within the domain of prosody itself, clusters of features, such as intonation, loudness, voice quality, etc., can be shown to work together in the accomplishment of specific actions. Very rarely do individual prosodic features signal anything by themselves. For example, Local et al. (1986) found for Tyneside English that turns that are treated by participants as complete typically involve a slowing down in speech rate towards the end of the turn, a sudden increase and decrease in loudness on the last stressed syllable, lengthening of that syllable, centralized quality on the final vowels and either a pitch step-up, or a drop in pitch on the final stressed syllable. Similarly, Local et al. (1985) describe the prosodic features of turn completion for London Jamaican English and Wells and Peppè (1996) for Ulster English. These publications show that conversationalists do not orient to single prosodic cues, such as intonation, but instead use and orient to combinations of prosodic and other practices in their negotiation over turn taking issues.

Empirical research findings on turn completion are in stark contrast to the prevailing assumption that turn and/or sentence completion is achieved by specific pitch patterns: low falling intonation for statements and WH-questions, high rising intonation for Yes/No questions. This assumption is present in many academic publications, even within the field of discourse and conversation analysis

[2]The British school of intonation has been pursued, for example, in O'Connor and Arnold (1961/1973), Halliday (1967, 1970), Crystal (1969), Cruttenden (1997), Wells (2006). Reviews of the approach can be found in Gibbon (1976), Crystal (1969) and Couper-Kuhlen (1986).

[3]Exceptions are the conversational phoneticians John Local and Richard Ogden, whose work is based in Firthian Prosodic Analysis (Ogden and Local 1994).

(Chafe 1980, 1987, 1993; Du Bois et al. 1993; Ford and Thompson 1996). For example, Chafe (1980: 20) claims:

> Even in listening to an unfamiliar language one can hear that every so often an idea unit ends with that distinctive falling intonation contour which we naturally associate with 'the end of a sentence'. To us English speakers, certainly, and probably to the speakers of most or all languages, this sentence-final intonation communicates an impression of completeness: the impression that the speaker has come to the end of something which has some kind of significant closure. This impression contrasts with the impression of incompleteness given by the intonation contours at the end of other idea units (…). From intonation alone, then, language sounds as if a series of nonfinal information units is punctuated every so often by some kind of finality.

The conceptual link between falling intonation and finality on the one hand and rising intonation and incompleteness on the other, is also a widespread claim in teaching materials for English pronunciation. For example, Hancock (2003) has a chapter entitled 'continuing and finishing tones' (p. 124–125), in which he establishes a connection between rising intonation as continuing and falling intonation as finishing:

> When we are telling someone a piece of news, we often check that they know the background to the story first. When we do this, the voice goes up at the end. Then, when we finally tell the news, the voice goes down at the end. This shows that we have finished the story. … If we are saying a list of things, our voice goes down at the end of the last thing to show we have finished. On the other things, the voice goes up to show the list is *not* finished. (2003: 124, emphasis in the original)

Similarly, Powell (1996: 41) teaches that 'a sharp fall gives weight and finality to what you have just said', while

> keeping your voice up tells the audience you are in the middle of saying something and mustn't be interrupted. Letting your voice drop lets them know you've completed what you wanted to say.

A final example comes from Dalton and Seidlhofer (1994: 58), who write:

> Intonation, or more precisely the factor of key or pitch height, is one important cue indicating a speaker's desire to continue his or her turn, or willingness to give it up. Non-low pitch is normally a signal for wanting to hold a turn, and low pitch for yielding it.

Regarding the intonation of questions, Cunningham and Bowler (1999, *New Headway Intermediate*) and Bowler and Cunningham (1999, *New Headway Upper-Intermediate*) also teach a clear link between form and function:

> In Yes/No questions, or in statements that are made into questions, the intonation normally goes **up** at the end. … In *Wh-* questions, the intonation normally goes **down** at the end. (Bowler and Cunningham 1999: 16, emphasis in the original)

In interactional linguistic research, however, this simplistic connection has not only been replaced by an awareness of clusters of phonetic and other features, but more recent studies have gone even further in deconstructing the connection between prosodic form and grammatical form/discourse function, particularly in terms of 'final intonation'. Instead, what is emerging from the latest enquiries into this area is that prosody must be studied in relation to social actions, rather than being linked to context-free notions of grammar, meaning and function. Two recent

investigations, Couper-Kuhlen (in press) and Szczepek Reed (2004), can be used here as an example.[4]

Couper-Kuhlen (in press) explores an issue that has been widely discussed by TESOL researchers and practitioners: the prevailing assumption that WH-questions end in falling and Yes/No questions in rising intonation. Researchers from various fields have long expressed severe doubts regarding this claim. For example, Fries (1964) analysed over 2,500 Yes/No questions from a television game show and found that 61.7% had falling intonation, whereas only 38.3% ended in rises (Fries 1964: 248).[5] Given these findings, the suggestion has been made that the difference between falls and rises should be linked to speaker knowledge instead. These distinctions are sometimes referred to as conducive *vs*. non-conducive questions. Conducive questions, i.e., those for which the questioner displays him/herself as knowing the answer, have been claimed to have falling intonation, whereas non-conducive questions are claimed to end in rises (Brown et al. 1980; Tench 1988; Thompson 1995).

In spite of these discussions, it is still a commonly held view throughout pronunciation teaching materials that in English, Yes/No questions end in rising and WH-questions in falling intonation. In an attempt to address this issue, Couper-Kuhlen (in press) presents a detailed interactional analysis of the local sequential structure of questions in a corpus of radio phone-in broadcasts. She, too, finds no straightforward link between final tones and grammatical question format. Instead, like Fries' (1964), her results show 'an overall numerical preference for falls over rises in conversational questions' (p. 12). Regarding individual question formats, Couper-Kuhlen finds many exceptions from perceived rules. For example, almost half of the Yes/No questions in her corpus end in falling intonation; and a quarter of WH-questions end in rises. WH-questions followed by final terms of address always end in rising tones, whereas almost all declarative forms used as questions end in falls.

However, Couper-Kuhlen does not claim that intonation is employed randomly. A closer analysis of the conversational actions achieved by individual question turns reveals interesting correlations. Distinguishing between four action types, topic proffers, topic follow-up questions, news receipts and newsmarks, and next-turn repair initiators, Couper-Kuhlen finds preferences for specific intonation patterns for individual actions. For example, participants show a preference for rises on topic proffers (63%) and next-turn repair initiators (62%); a slight preference for falling intonation on follow-up questions (55%), and a clear preference for falls on news receipts and newsmarks (65%). However, these findings are tendencies

[4]In addition to those mentioned here, one of the most prevailing form-function associations occurs in Brazil's 'proclaiming' (fall and rise-fall) and 'referring' (rise and fall-rise) tones, which are said to indicate new and given information, respectively (Brazil et al. 1980; Brazil 1985).

[5]In a related study, Levis (1999) conducted listening experiments to investigate the distinction between low-rising and high-rising pitch on Yes/No questions in American English. Challenging the assumption that American English Yes/No questions end in a high rising tone, with low rising intonation being used only in British English, Levis shows that his American English speaking subjects in fact do not differentiate between the two contours.

rather than predictable patterns. It is only when the variables conversational action and grammatical format are brought together that there are some clearly observable distributions. For example, for topic proffers in the Yes/No question format, rises are clearly the preferred option (81%). Topic follow-up questions in the WH-format strongly prefer falling intonation (79%). And for news receipts and newsmarks in the repeat question form, rises clearly dominate (91%). As a result of these findings, Couper-Kuhlen concludes that

> meaningful generalizations about final intonation in questions can only be made within conversational activity types and together with specific syntactic question forms. There are no consistent patterns for final intonation either across conversational activities irrespective of syntactic type, or across syntactic types irrespective of conversational activity. Because there is also a significant skewing of syntactic question types across conversational activity, much speaks in favor of focusing first on the ACTION being implemented by a spate of talk, and then on its linguistic and prosodic form. Once a particular conversational activity has been singled out for attention, it can be meaningfully asked (i) what syntactic form and (ii) what intonational or prosodic formatting is typically deployed for the task under consideration. (p.15, emphasis in the original).

In a subsequent part of her study, Couper-Kuhlen goes on to make a convincing argument that for Yes/No questions the local epistemic context is crucial. She shows that varying degrees of displayed speaker certainty are closely linked to choices over rising or falling tones. Questions that are designed as displaying epistemic uncertainty frequently end in rising intonation; whereas questions that display greater certainty often co-occur with falling tones.

Both strands of findings from Couper-Kuhlen's study support previous claims. There is no direct link between grammatical question format and final pitch contour; and for Yes/No questions, displayed speaker knowledge seems to play an important role. Furthermore, Couper-Kuhlen's analysis reveals a link between conversational actions, final pitch and grammatical format – a combination which is perhaps not surprising, given that language from a conversation analytic perspective is a set of simultaneously employed resources for social interaction.

Besides the differences between question intonation in real-life conversation and in some pronunciation teaching materials, there are additional discrepancies between teaching practice and Couper-Kuhlen's findings. For example, the preferred falling pitch patterns she describes for follow-up questions are not in accordance with Cunningham and Bowler (1999), who teach that 'to show that you are interested and want to hear more, your intonation should start high, go down and then go up at the end' (p. 17). Couper-Kuhlen's finding regarding the role of prosody for the display of epistemic certainty is in contrast to the pattern taught by both Hancock (2003) and Hewings (2007), in which 'finding out information that we don't already know' is ascribed falling intonation; and 'making sure that information we think we know is, in fact, correct' is said to have rising final pitch (Hewings 2007: 92).

Throughout her study, Couper-Kuhlen makes clear that an analysis of intonation in natural talk must always take into consideration participants' local interactional environment. Broad generalizations concerning intonational functions are questionable, at best. However, her work does suggest a combined significance in clusters of

prosodic, grammatical and action formats. In the second half of this section we turn away from the specific prosodic format of questions, to the role of prosody for turn taking as such. As we approach the large variety of interactional projects that are implemented through turns besides questions we find that the connection between prosody and turn completion becomes even less tangible.

Regarding prosody and turn taking, publications on English pronunciation typically focus on the difference between falling and rising tones and their function for turn completion or continuation. The assumption prevails that we can establish rules regarding prosodic design and the projection of how turns will end, irrespective of the actions being accomplished through those turns. For example, Jenkins (2004: 111) writes that some materials have 'tended to … restrict [the subtleties of discourse-based tone assignment] mainly to matters of conversation management (turn-taking, etc.) where … it is easier to articulate "rules"'. However, careful analysis of natural talk soon reveals two basic facts. First, prosody plays a role in the implementation of many actions, not only turn projection; and second, turn projection is accomplished by additional interactional cues besides prosody. This means that in designing turns for their recipients, participants do more than signal places of potential transition or continuation; and when they do so, this involves clusters of prosodic and other cues. Turns-at-talk are the way by which interactants implement and coordinate actions. Where those actions start and finish is only one, however important, part of the equation. Given the wide variety of social actions accomplished through talk, it is not surprising that the role of prosody is a multi-layered one.

With these considerations in mind, Szczepek Reed (2004) investigated turn final intonation in English more generally. Taking as a starting point the above mentioned widespread assumption that turn completion in English is achieved either by falls-to-low or rises-to-high, and that other intonation patterns signal incompletion, two corpora of spontaneous American English interactions were analysed. While the data did contain turn transitions following the two aforementioned patterns, they also showed a large number of smoothly accomplished turn transitions after other pitch contours. Four further intonation patterns occurred regularly in turn final position: level intonation, rise-to-mid intonation, pitch step-ups and stylized intonation. For example, in the following extract Lynne and Lenore are talking about their plans for the day. Lenore's turn *I don't know* at line 12 ends in level pitch. Transcription notations can be found in the Appendix.

(1) SBC001 Actual blacksmithing (Szczepek Reed 2004: 105)

1	LY:	↑I don't know what her PLANS really are;
2	(0.09)	
3		but i thInk pretty much just gO Out;
4		and take cAre of 'em and then;
5		.hh
6		maybe gO to that-
7	(0.49)	
8		sEAsonal DANCE or whatEver it is,
9	(1.9)	

10		<<h>is THAT what it's CALLED?>
11	(0.9)	
12	**LE:**	**I don't knOw -**
13	(0.7)	
14	LY:	<<all+p>don't know what's>cAlled Either;

Level intonation is not a pitch pattern that has previously been recognized as typical for turn completion. In this extract, however, the participants treat it as a default practice and listening to the extract reveals nothing out of the ordinary. Lenore's response turn *I don't know* is clearly designed as complete, which is evidenced by the absence of any further talk from her and by Lynne's non-competitive next incoming (line 14).

The following extract shows an example of turn final rise-to-mid intonation, another pattern traditionally assumed to be a cue for turn continuation. Pamela and Darryl are discussing life after death. Darryl's reply *I've read that* at line 8 ends in a rise-to-mid.

(2) SBC005 A book about death (Szczepek Reed 2004: 109)

1	PA:	pEOple who (0.08) hA:d (1.04) TECHnically dIEd;=
2		and then had been reVIVED;
3	(0.83)	
4		SAW;
5	(0.17)	
6		**RELatives COMing for them –**
7	(0.28)	
8	**DA:**	**I've READ thAt,**
9	(0.62)	
10	PA:	.hh
11		**cOUrse thAt may be what happens: –**
12	(0.28)	
13		prior to the BIG^{h} –
14	(1.71)	
15		the big NOthing.

Darryl's turn *I've read that* is clearly designed as complete and is treated as such by Pamela. Darryl does not continue speaking after the final syllable and Pamela comes in to speak without signs of turn competition. The intonation on the final part of Darryl's turn is a slight rise in pitch, represented by the comma in the transcript. Note also that the immediately previous turn transition occurs after a final level tone (line 6).

It is important to note in this context that Szczepek Reed (2004) does not argue for either level or rise-to-mid pitch as turn completion cues. It is clear even from within the above extracts that at other turn locations the same pitch accents accompany non-final words. For example, line 11 in extract (2) shows level pitch co-occurring with a spate of talk that in another context could be designed as complete; however, here it is clearly neither designed nor treated as such. What is being argued is

that pitch is used as a cue for a variety of actions, implemented through turns. In the extracts above it is the nature of the conversational action that determines participants' choice of intonation; turn projection seems to be smoothly accomplished as part of that action, and with the specific prosodic cues employed at this sequential location.

Both Couper-Kuhlen (in press) and the study presented above show that interaction is too complex a negotiation process to be described or taught, according to simplistic form-function relations. Further evidence for this claim comes from recent research by Marianna Kaimaki (2011). Her analysis of news receipts in natural English conversation suggests that for non-valenced news receipts such as *oh really* final rising and falling intonation are in free variation, but with certain preferences for the interactional design of next turns. For example, *oh really* with rising intonation has a tendency to be followed by confirmation and subsequent elaboration by the next speaker, whereas the same token with falling intonation is frequently followed by talk from the producer of the token. Along a similar vein, Flax et al. (1991) were not able to relate specific prosodic forms to communicative functions in their analysis of three mother-child interactions. Furthermore, conversational experiments by de Ruiter et al. (2006) suggest that intonation patterns play little or no role in allowing listeners to predict upcoming utterance completion points.

The findings reported here do not suggest that participants use intonation patterns entirely at random. However, as these studies demonstrate, during talk participants are engaged in several domains of action simultaneously. Turn taking is only one of these domains. In English, prosody is also employed as a cue for other conversational actions, such as assessments, repair, reported speech, marking new sequential beginnings, to name just a few. Furthermore, there are prosodic distinctions made in the domains of phonetics, such as the durational difference between long and short vowels; lexis, such as the pitch, loudness and durational differences necessary for word stress; and syntax, such as the role of pauses for separating one syntactic construction from another. Moreover, prosody is used as a cue for the display of attitude, affect and stance. Many of these distinctions occur simultaneously at any point in conversation, making it impossible to link specific prosodic events with individual linguistic functions.

In addition, most interactional activities, such as turn taking, are not accomplished through prosody alone. Instead, prosody is one of many resources that participants employ in their negotiations over actions and meaning. Other resources include articulation, lexis, grammar, information structure and, if the interaction is face-to-face, gesture, gaze and body posture, to name only the most obvious. In summary, conversational prosody is used with a large number of interactional requirements and actions, often simultaneously; but rarely are any of those requirements and actions accomplished exclusively through prosody. This means that in teaching English pronunciation, at least *conversational* pronunciation, teaching a close form-function relation between prosody and single discourse domains neither reflects real-life conversational practice, nor enables students to accomplish conversational actions successfully.

10.3 Prosody and Interactional Alignment

In this section we move away from specific prosodic practices, such as rising or falling intonation, and proceed to the overall prosodic shape of turns-at-talk. While it was argued in the previous section that individual tones cannot be linked to single interactional functions, this section shows that for the accomplishment of one of the most basic interactional tasks, the prosodic format, while playing a defining role, cannot be predicted out of context. Instead, it can only be specified in the local environment of its sequential location and as developing from the immediately preceding conversational actions.

In analysing conversation, it is of primary importance to remember that, for participants, interaction unfolds step-by-step and in real time. For conversationalists, the emerging interactional development is never predictable and each spate of talk is followed by contingencies, rather than certainties. As analysts, however, we have the luxury of seeing after the fact and at a single glance how actions, or patterns, are collaboratively accomplished by participants. Given the emergent and negotiating nature of spontaneous interaction, it is not surprising that linguistic and other interactional strategies show a high degree of flexibility, allowing participants to adapt to continuously newly arising situations. Of the interactional signalling domains participants have at their disposal, prosody has been shown to be one of the most locally adaptive. Auer (1996), for example, demonstrates that while syntax operates on a more global interactional scale, prosody is used for immediate, local decisions.

One of the most fundamental, most local and most frequent decisions participants must make in the course of a conversation is whether the turn they are about to produce is going to be designed as a continuation of a previously established action trajectory, such as an answer to a question or a return greeting to a greeting; or whether it is going to start a new trajectory, such as a repair initiation following a question instead of an answer, or a *how-are-you* following a return greeting. Designing turns as either FIRSTS or SECONDS is a basic interactional requirement that newly emerges at every turn transition.

In connection with this fundamental interactional necessity, prosody offers an equally fundamental format. This format has been referred to as *prosodic orientation* (Szczepek Reed 2006, 2009a, b, 2010b). The term describes participants' prosodic display of awareness of other speakers' prosody. One form of prosodic orientation is *prosodic matching*, whereby a next speaker employs the same prosodic pattern as an immediately prior speaker. Studies of prosodic orientation have shown that it is a fundamental interactional practice by which participants align next turns with prior talk. Prosodic orientation typically co-occurs with acknowledgements, answers to questions, seconds in opening and closing sequences and other turns in next position (Szczepek Reed 2006: 65–88).

Earlier research on conversational prosody already revealed that in certain interactional contexts the practice of prosodic matching is more significant interactionally than any individually defined prosodic format. Couper-Kuhlen (1996) shows that participants accomplish different actions, depending on whether they match a previous

speaker's pitch register on an absolute or relative scale. In analysing interactions between a male radio host and his female callers in a radio phone-in programme, she finds that when the host repeats his callers' high pitched turns with a pitch register that is also high, but relatively so within his own voice range, this is treated as default repetition. In contrast, when he repeats callers' turns with perfect pitch register matching, thus speaking extremely high in his own voice range, this is treated as mimicry and therefore, as implicit criticism. Couper-Kuhlen's work shows that in these specific circumstances, the prosodic format is not interpreted according to its monologous, context-free design but in relation to the prosody used by previously speaking participants.

Szczepek Reed (2009a, b) demonstrates that prosodic orientation is a primary device for designing talk as a second, i.e., as a turn responding or aligning with a prior turn, rather than beginning a new action trajectory. Simultaneously, those studies show that refraining from prosodic orientation is a way of designing talk as the beginning of something new. An analysis of British and American English radio phone-in programmes revealed that when callers first come on the air they have two choices. They can either design their talk as a continuation of the introductory turn by the radio host, who has typically introduced them and welcomed them to the programme; or they can design their talk as a new sequential beginning, without responding to the host's introduction. In a collection of 131 opening sequences, callers consistently designed seconds as prosodically orienting and firsts as using a new, non-orienting prosodic format. This even held if the lexical items they were using were similar: if the items were not prosodically orienting to prior talk, they were not treated as seconds by participants. See, for example, the following two opening sequences. In both cases, host Dave greets his callers with a greeting token and in both cases, callers produce a greeting token in their next turn. However, only in one of the cases is that second greeting token actually treated as a return greeting.

(3) Brainteaser: Nigel2 (Szczepek Reed 2009a: 1229)

1	DA:	next is NIgel HIBbits;
2		who lives in PRESTwich.
3		**<<h>↑HI `NI:GE,>**
4	**NI:**	**<<h>↑HI `DA:VE,>**
5	DA:	<<all>how ARE ya.>
6	(0.25)	
7	NI:	.hh
8		nOt too BAD,
9	DA:	GOOD to speak to you agAIn,

(4) Brainteaser: Ann (Szczepek Reed 2009a: 1229)

1	DA:	a:nd we have ANN,
2		who lives in GORton.
3	(0.23)	
4		who's FIRST.=
5		and then of COURSE,

6		.h
7		After our two callers we do have RAchel back again.
8		.h
9		**ANN.**
10		**HI.**
11	(0.26)	
12	**AN:**	**<<breathy>HELL: ʼO:.>**
13	**DA:**	**<<breathy>HELL: ʼO:.>**
14		<<h>how ARE you Ann,>
15	(0.25)	
16	AN:	I'm FINE,
17		THANKS,
18	DA:	GOOD.
19		WELcome to piccadilly rAdio.

In extract (3), the radio presenter's introductory turn *hi nige* (line 3) is designed by him and treated by caller Nigel as a first greeting. Dave's introductory greeting is produced with a high pitch onset, high overall pitch register and falling-rising intonation. The address term *nige* is delivered with lengthening and both syllables carry equal primary stress. Nigel's return greeting contains exactly the same prosodic features: high pitch onset, high pitch register, falling-rising intonation, lengthening on the monosyllabic address term and primary stress on both syllables. Following Nigel's turn, Dave produces a new first, *how are ya* (line 5). This shows that he is treating Nigel's turn as a second, i.e., as a return greeting to his greeting, which warrants moving on to the next sequential activity.

In contrast, extract (4) shows a different pattern at work. Presenter Dave once again produces a greeting token and an address term (line 9–10), and his caller Ann also produces a greeting token in her next turn (line 12). Reading the transcript, these two turns could be interpreted as a greeting pair. However, following Ann's greeting, Dave produces another greeting token (line 13). This shows that he is not treating Ann's previous greeting as a return greeting to his earlier turn but as a new first, requiring a second. The prosodic format of this sequence is noticeably different from the previous one. Ann's turn following Dave's introduction displays no prosodic resemblance to his. Instead, it is delivered with breathy voice quality, sound and syllable lengthening and rising-falling intonation on the last syllable. As Dave comes in with his third greeting token, he displays the same prosodic format: breathy voice quality, lengthening and rising-falling final pitch. Thus, Ann's lack of prosodic orientation plays an important role in designing her turn as a new first, rather than a second; while Dave's subsequent prosodic orientation designs his return greeting as a second to Ann's first.

Examples such as those above clearly show that it is not individual prosodic patterns, such as certain types of intonation contour, stress pattern, pitch register or voice quality, that make for turns to be interpreted by their recipients as either responses or new beginnings. Instead, the decisive factor is whether they repeat the prosodic design of the immediately prior turn, or not.

This argument is further supported by research currently being conducted by Gorisch et al. (2010), who are investigating the use of *uhu* as a continuer on the one hand and as an acknowledgement token on the other. Interestingly, 'no difference in patterns of prosody (F0-range, duration, F0 movement), and gesture (smiles, nods, blinks) or gaze (to speaker, mid-distance gaze) could be found between tokens with the two different conversational functions.' Instead, what the authors find is that in the function of continuer, participants design *uhu* as a prosodic copy of the end of the previous turn. Gorisch et al. conclude that it is this practice of prosodic matching of prior talk, rather than any specific prosodic, or other, practices in and of themselves, that determines the interactional function of *uhu*.

The practice of prosodic matching has also been linked to interactional, rather than sequential alignment. Wells and Corrin (2004) and Tarplee (1996) show that prosodic matching is a frequent practice in carer-child interactions, where it works as a collaborative, affirming practice. Skidmore (2008) and Roth and Tobin (2010) make a similar observation for teacher-student interactions the classroom. Furthermore, prosodic alignment of various forms frequently occurs during collaborative turn sequences, during which it is part of an interactional strategy to design a next speaker's contribution as being part of a turn begun by an immediately prior speaker. See, for example, the following extract, in which incoming speaker Barbara completes a turn by previous speaker Patrick (lines 7–9).

(5) 29: Rubbish

1	PA:	but you CA:N use quality meat [for SAUSages.
2	BA:	[VEAL actually,
3	RO:	Oh you no you you CA:N,
4		and and they DO:,
5		[in in GERmany ↑And swItzerland,
6	PA:	[but the but the ma↑JOrity of sAUsage:s,
7		**A::RE,**
8		[()
9	**BA:**	**[↑RUbbish.**

Barbara's prosodic design is such that it completes Patrick's. His turn emerges as two intonation phrases ending in final rises with lengthening (lines 6, 7), which can be heard as preparing for a final pitch peak (Schegloff 1998). This pitch peak is produced by Barbara in her collaborative completion *rubbish* (line 9). By completing previous participants' prosodic designs, next speakers can embed their utterances into theirs and thus share turns-at-talk. This type of prosodic backwards orientation is both a form of sequential alignment, in that it accomplishes a specific turn design, and a form of interactional alignment, in that it displays agreement with a previous speaker. As in previous examples, the prosodic format of the incoming utterance is not meaningful in and of itself, but only in direct relation to the prosodic design of prior talk.

10.4 Implications for Teaching English Pronunciation

It is clear from the interactional linguistic research presented above that language-in-conversation is not appropriately defined as a tool for individual actions alone, but must be approached as a resource and negotiating strategy for social interaction. Prosody, therefore, must be described according to its role for both the accomplishment and the coordination of actions across turns and participants. In order to teach prosody against such a background, one would want to introduce a perspective on language and linguistic meaning that is, first, based entirely on real-life discourse practices; and second, rooted in participants' mutual and collaborative orientation to each other. The first aspect is not at all new; in fact Communicative Language Teaching is the prevalent teaching method for ELT and TESOL practitioners today, at least in the majority of English speaking countries (Cf. Widdowson 1978; Littlewood 1981/2007; Savignon 1991; Richards 2005; Spada 2007). However, while the demand to teach language as communication rather than cognitive activity has brought about an important change in teaching practice, the approach nevertheless falls short of a full appreciation of talk as spontaneously emerging collaborative negotiation and coordination of actions.

Fundamentally, talk-in-interaction involves (at least) two participants. In any language and whatever the degree of linguistic competence, those participants collaborate with each other in their accomplishment of social actions. No sound/word/sentence/turn is interactionally meaningful without interactional work by others, which may take the form of, minimally, displayed recipiency and maximally, displayed uptake. Non-native speaker interactions, just as interactions between native speakers, involve continuous negotiations over action, sequential structure and meaning. Recent cross-cultural research in CA suggests that the broad interactional projects that are pursued in conversation display marked similarities across languages, even if the practices deployed for their pursuit may vary (cf. Sidnell 2009; Enfield et al. 2010). Among the activities all participants in spontaneous conversation are continuously and simultaneously involved in are those listed by Schegloff et al. (2002: 4–5):

> Whether speaking their native language or another, whether fluently or not, whether to another or others doing the same or not, whether in ordinary conversation or in a classroom or in the work place or in some other institutionally or functionally specialized situation, there are certain issues all participants in talk-in-interaction will find themselves dealing with. They will, for example, need some way of organizing the order of their participation … (turn-taking). They will fashion their contributions to be recognizable as some unit of participation … (turn organization). They will have practices for forming their talk so as to accomplish one or more recognizable actions (action formation). They will deploy resources for making the succession of contributions cohere somehow, either topically or by contributing to the realization of a trajectory of action or interaction (sequence organization). They will avail themselves of practices for dealing with problems in speaking, hearing and/or understanding the talk (organization of repair). They will select and deploy and understand the words used to compose the talk, and will do that in a timely fashion (word/usage selection). They will do all of this with an eye to their co-participants (recipient design) and to the occasion and context, its normative parameters or boundaries of duration, appropriate activities and their order, etc., (overall structural organization of the occasion of interaction).

In many if not most interactional encounters, there exists a degree of variation in the practices used for accomplishments such as those mentioned above. For example, speakers of the same native language may come across an intonation contour they have not previously heard being used in a certain manner, due to the linguistic variety spoken by their co-participants. This does not routinely lead to a breakdown in communication but simply to more negotiation. In the same way, non-native speaking participants and their native or non-native speaking co-participants employ negotiation strategies in order to achieve conversational actions successfully. Were we to place the burden of responsibility for successful communication squarely on non-native speakers by demanding of them 'correct' prosodic patterns, we would forget that in native-native talk this burden is always shared.

Lindemann (2006) issues a strong call for the role and involvement of native speaking interlocutors in interactions with non-native speakers. Furthermore, the research conducted by Barbara Seidlhofer and her students and colleagues on English as a Lingua Franca shows how successful non-native speakers are at achieving interactional outcomes, in spite of deviations from native speaker practices (Seidlhofer 2004: 220). Given these convincing voices and given the high degree of complexity involved in relating prosody and interaction, of which the above sections have only scratched the surface, I strongly agree with Jenkins (2000) that above and beyond basic intelligibility, the decision to aim for specific, native-like pronunciation patterns should be left up to learners.

If prosodic patterns are to be learnt and taught, interactional linguistic research does not support a teaching practice of prosodic formats in isolation. As has been pointed out above, conversational actions are rarely accomplished by individual prosodic cues alone; and prosody is simultaneously deployed for a number of interactional modes. The approach to pronunciation teaching put forward by Seidlhofer and Dalton-Puffer (1995) takes this into consideration. They suggest that 'prefabricated chunks' (1995: 135), and 'intonational idioms' (Dalton and Seidlhofer 1994: 45), based on the phonological category of the 'tone unit' (Cf. Crystal 1969; Cruttenden 1997; Wells (2006)), make available a holistic learning experience. However, tone units in their traditionally defined form cannot easily be applied to natural talk (Szczepek Reed 2010c, d); a concession readily made by those authors (Dalton and Seidlhofer 1994: 47). While spontaneous speech is indeed divided by participants into shorter spates of talk, these spates do not typically display the prosodic and grammatical features ascribed to tone units in traditional phonology. However, Szczepek Reed (forthcoming) shows that chunks of talk are frequently employed by participants as 'action components'. Therefore, a teaching practice that introduces prosody as one aspect of lexical phrases, or chunks, may be a first step. One that takes into consideration potentially emerging sequential locations and actions for such phrases would go even further in marrying up real-life pronunciation with its acquisition.

However, because conversation analytic descriptions of recurring interrelations between prosodic and other cues are based entirely on their local placement, applications of specific research findings will always have to involve a heightened awareness of context by both teacher and learner. Schegloff et al.'s (2002: 18) warning regarding applications of CA research in general is also relevant for

prosodic features in particular. 'Specific findings should not be used to categorize talk in other settings without investigating whether similar practices are used to accomplish similar actions in the new setting.' Furthermore, individual prosodic practices should be considered in their interactional context by teachers and examiners. For example, a perspective on silence as 'hesitation', and thus as a form of a linguistic error would miss the fact that pauses play an important role, both for prosodic cues such as speech rate and rhythm (Couper-Kuhlen 1993), but also for the implementation of specific actions (Szczepek Reed 2009a). Teaching methods for pronunciation would ideally expose learners to these variations, while oral assessment practices would preferably take them into consideration as natural aspects of spontaneous speech (cf. McCarthy 2009).

From a conversation analytic perspective, teaching and assessment methods that help improve learners' interactional negotiating skills are generally to be preferred over those that aim at purely cognitive skills through drills and pattern practice (Cf. Wagner 1996; Firth and Wagner 1997). Concerning specific teaching methods, highly communicative approaches such as Cooperative Language Learning (Cf. Kessler 1992; McCafferty et al. 2006) and Task-Based Language Teaching (Cf. Nunan 2004; van den Branden et al. 2009) seem to provide a maximum amount of opportunities for spontaneous interaction. Most importantly, learners would ideally be given the chance to practice interaction with a raised awareness of the seemingly universal requirements of turn design, turn taking, sequence organization, action formation and repair. Crucially, these requirements are only ever met successfully by (native or non-native speaking) participants through continuous monitoring of their co-participants' talk, including their prosody. Therefore, learners of English must be alerted to the collaborative nature of prosody, so as to enable them to place their prosodic delivery in relation to that deployed by others.

A strong emphasis on orientation, and display of orientation, to others would require an increased focus on listening skills – not only in the sense of listening comprehension but primarily, in the sense of being attuned to the (prosodic) behaviour of others at all times. It is this orientation to co-participants that is at the heart of conversation and of conversational prosody, allowing participants to engage in, or refrain from, sequential and interactional alignment. Therefore, one of the most significant teaching goals for pronunciation and conversational skills in general, is the teaching of ongoing orientation to other speakers.

Such a focus would mean a re-conceptualization of some central speaking skills, one being fluency. While defined in a variety of ways, such as general proficiency, rapid speech rate or naturalness (cf. Chambers 1998), fluency is typically treated as the result of monologic performance. With regard to the collaborative nature of talk and in particular, the interactionally achieved nature of many prosodic functions, a more appropriate way of approaching fluency is the one put forward by McCarthy (2009), who suggests viewing fluency as an 'interactive achievement',

> …perhaps more adequately captured by the metaphor of confluence. Achieving confluence, successfully interacting in talk that flows and being perceived as both able to create within one's own utterances and across utterances the satisfactory perception of flow for all participants is an art, the evidence of which will not be found or fairly assessed in monologic contexts but in the robust evidence of dyadic and multi-party talk. (McCarthy 2009: 23)

A long-term goal for a conversation analytically inspired approach to English language teaching would be to view not only fluency but language proficiency and even language as such as a collaborative accomplishment by participants, native or non-native. Such a perspective would open up the opportunity for methods to treat *social interaction through language* as the entity to be taught and tested. In pursuit of this goal, this chapter has argued for applying the fundamental basis of talk-in-interaction, that is, collaborative negotiation emerging through real time, to pronunciation teaching methods. Prosody, in particular, is an interactional resource for bridging the gap between successive turns and actions. To enable students to use prosody for such collaboration and bridging work is perhaps one of the highest achievements of any pronunciation teaching methodology.

Appendix

Transcription Conventions (Adapted from Selting et al. 1998)

Pauses and lengthening	
(2.85)	measured pause
:::	lengthening
Accents	
ACcent	primary pitch accent
Accent	secondary pitch accent
Phrase-final pitch movements	
?	rise-to-high
,	rise-to-mid
-	level
;	fall-to-mid
.	fall-to-low
Pitch step-up/step down	
↑	pitch step-up
↓	pitch step-down
Change of pitch register	
<<l> >	low pitch register
<<h> >	high pitch register
Volume and tempo changes	
<<f> >	forte
<<p> >	piano
<<all> >	allegro
<<len> >	lento
Breathing	
.h, .hh, .hhh	in-breath
h, hh, hhh	out-breath
Other conventions	
[	overlapping talk
[	

References

Auer, P. 1996. On the prosody and syntax of turn-continuations. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 57–101. Cambridge: Cambridge University Press.

Barth-Weingarten, D., A. Wichman, and N. Dehé, eds. 2009. *Where prosody meets pragmatics*. Bingley: Emerald.

Barth-Weingarten, D., E. Reber, and M. Selting, eds. 2010. *Prosody in interaction*. Amsterdam: Benjamins.

Betz, E. 2008. *Grammar and interaction: Pivots in German conversation*. Amsterdam: Benjamins.

Bowler, B., and S. Cunningham. 1999. *New headway pronunciation course: Upper-intermediate*. Oxford: Oxford University Press.

Bowles, H., and P. Seedhouse, eds. 2007. *Conversation analysis and language for specific purposes.* Bern: Peter Lang.

Brazil, D. 1985. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Brazil, D., M. Coulthard, and C. Johns. 1980. *Discourse intonation and language teaching*. London: Longman.

Brown, G., K. Currie, and J. Kenworthy. 1980. *Questions of intonation*. London: Croom Helm.

Bybee, J., and M. Noonan, eds. 2001. *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*. Amsterdam: Benjamins.

Chafe, W. 1980. The deployment of consciousness in the production of a narrative. In *The pear stories: Cognitive, cultural and linguistic aspects of narrative production*, ed. W. Chafe, 9–50. Norwood: Ablex.

Chafe, W. 1987. Cognitive constraints on information flow. In *Coherence and grounding in discourse*, ed. R. Tomlin, 21–51. Amsterdam: Benjamins.

Chafe, W. 1993. Prosodic and functional units of language. In *Talking data: Transcription and coding in discourse research*, ed. J.A. Edwards and M.D. Lampert, 33–43. London: Lawrence Erlbaum.

Chambers, F. 1998. What do we mean by fluency? *System* 25(4): 535–544.

Couper-Kuhlen, E. 1986. *An introduction to English prosody*. Tübingen: Max Niemeyer.

Couper-Kuhlen, E. 1991. A rhythm-based metric for turn-taking. In *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 1, 275–278. Aix-en-Provence: Service des Publications, Université de Provence.

Couper-Kuhlen, E. 1993. *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam: Benjamins.

Couper-Kuhlen, E. 1996. The prosody of repetition: On quoting and mimicry. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 366–405. Cambridge: Cambridge University Press.

Couper-Kuhlen, E. 2004. Prosody and sequence organization: The case of new beginnings. In *Sound patterns in interaction*, ed. E. Couper-Kuhlen and C.E. Ford, 335–376. Amsterdam: Benjamins.

Couper-Kuhlen, E. in press. Some truths and untruths about final intonation in conversational questions. In *Questions. formal, functional and interactional perspectives*, ed. J.P. de Ruiter. Cambridge: Cambridge University Press.

Couper-Kuhlen, E., and C.E. Ford, eds. 2004. *Sound patterns in interaction.* Amsterdam: Benjamins.

Couper-Kuhlen, E., and B. Kortmann, eds. 2002. *Cause, condition, concession, contrast: Cognitive and discourse perspectives.* Berlin: Mouton de Gruyter.

Couper-Kuhlen, E., and M. Selting, eds. 1996. *Prosody in conversation.* Cambridge: Cambridge University Press.

Cruttenden, A. 1997. *Intonation*. Cambridge: Cambridge University Press.

Crystal, D. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
Cunningham, S., and B. Bowler. 1999. *New headway pronunciation course: Intermediate*. Oxford: Oxford University Press.
Dalton, C., and B. Seidlhofer. 1994. *Pronunciation*. Oxford: Oxford University Press.
de Ruiter, J.P., H. Mitterer, and N.J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82(3): 515–536.
Du Bois, J.W., S. Schuetze-Coburn, S. Cumming, and D. Paolino. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, ed. J.A. Edwards and M.D. Lampert, 45–89. London: Lawrence Erlbaum.
Enfield, N.J., T. Stivers, and S.C. Levinson. 2010. Question-response sequences in conversation across ten languages. Special issue. *Journal of Pragmatics* 42(10).
Firth, A., and J. Wagner. 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal* 81: 285–300.
Flax, J., M. Lahey, K. Harris, and A. Boothroyd. 1991. Relations between prosodic variables and communicative functions. *Journal of Child Language* 18: 3–19.
Ford, C.E. 1993. *Grammar in interaction: Adverbial clauses in American English conversations*. Cambridge: Cambridge University Press.
Ford, C.E., and S.A. Thompson. 1996. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In *Interaction and grammar*, ed. E. Ochs, E.A. Schegloff, and S.A. Thompson, 134–184. Cambridge: Cambridge University Press.
Fox, B.A. 2001. An exploration of prosody and turn projection in English conversation. In *Studies in interactional linguistics*, ed. M. Selting and E. Couper-Kuhlen, 287–315. Amsterdam: Benjamins.
Freese, J., and D.W. Maynard. 1998. Prosodic features of bad news and good news in conversation. *Language in Society* 27: 195–220.
French, P., and J. Local. 1986. Prosodic features and the management of interruptions. In *Intonation in discourse*, ed. C. Johns-Lewis, 157–180. London: Croom Helm.
Fries, C.C. 1964. On the intonation of yes-no questions in English. In *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*, ed. D. Abercrombie, D.B. Fry, P.A.D. McCarthy, N.C. Scott, and J.L.M. Trim, 242–254. London: Longman.
Gibbon, D. 1976. *Perspectives of intonation analysis*. Frankfurt am Main: Peter Lang.
Golato, A. 2005. *Compliments and compliment responses: Grammatical structure and sequential organization*. Amsterdam: Benjamins.
Gorisch, J., B. Wells, and G.J. Brown. 2010. The response token as a copy of the prior turn: Prosodic and visual properties of ‚uhu'. Paper presented at the International Conference on Conversation Analysis (ICCA10), Mannheim, Germany, 4–8 July 2010.
Hall, J.K., J.A. Hellermann, and S. Pekarek Doehler. 2011. *L2 Interactional competence and development*. Clevedon: Multilingual Matters.
Halliday, M.A.K. 1967. *Intonation and grammar in British English*. The Hague: Mouton.
Halliday, M.A.K. 1970. *A course in spoken English: Intonation*. Oxford: Oxford University Press.
Hancock, M. 2003. *English pronunciation in use*. Cambridge: Cambridge University Press.
Hewings, M. 2007. *English pronunciation in use: Advanced*. Cambridge: Cambridge University Press.
Hutchby, I., and R. Wooffitt. 2008. *Conversation analysis*, 2nd ed. Cambridge: Polity Press.
Jenkins, J. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.
Jenkins, J. 2004. Research in teaching pronunciation and intonation. *Annual Review of Applied Linguistics* 24: 109–125.
Kaimaki, M. 2011. Sequentially determined function of pitch contours: The case of English news receipts. In *York papers in linguistics* (*Series 2*), Issue 11, 49–73.

Kelly, J., and J. Local. 1989. *Doing phonology: Observing, recording, interpreting*. Manchester: Manchester University Press.
Kessler, C., ed. 1992. *Cooperative language learning: A teacher's resource book.* New York: Prentice Hall.
Klewitz, G., and E. Couper-Kuhlen. 1999. Quote-unquote: The role of prosody in the contextualization of reported speech sequences. *Pragmatics* 9(4): 459–485.
Levis, J.M. 1999. The intonation and meaning of normal yes/no questions. *World Englishes* 18(3): 373–380.
Liddicoat, A.J. 2007. *An introduction to conversation analysis*. London: Continuum.
Lindemann, S. 2006. What the other half gives: The interlocutor's role in non-native speaker performance. In *Spoken English, TESOL and applied linguistics: Challenges for theory and practice*, ed. R. Hughes, 23–49. Basingstoke: Palgrave Macmillan.
Littlewood, W. 1981/2007. *Communicative language teaching*. Cambridge: Cambridge University Press.
Local, J. 1992. Continuing and restarting. In *The contextualization of language*, ed. P. Auer and A. di Luzio, 273–296. Amsterdam: Benjamins.
Local, J. 1996. Conversational phonetics: Some aspects of news receipts in everyday talk. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 177–230. Cambridge: Cambridge University Press.
Local, J., and J. Kelly. 1986. Projection and 'silences': Notes on phonetic and conversational structure. *Human Studies* 9: 185–204.
Local, J., B. Wells, and M. Sebba. 1985. Phonology for conversation: Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics* 9: 309–330.
Local, J., J. Kelly, and B. Wells. 1986. Towards a phonology of conversation: Turn-taking in Tyneside English. *Journal of Linguistics* 22: 411–437.
McCafferty, S.G., G.M. Jacobs, and A.C. DaSilva Iddings, eds. 2006. *Cooperative learning and second language teaching.* Cambridge: Cambridge University Press.
McCarthy, M. 2009. Rethinking spoken fluency. *Estudios de lingüística inglesa aplicada* 9: 11–29.
Nunan, D. 2004. *Task-based language teaching*. Cambridge: Cambridge University Press.
O'Connor, J.D., and G.F. Arnold. 1961/1973. *Intonation of colloquial English*. London: Longman.
Ochs, E., E.A. Schegloff, and S.A. Thompson, eds. 1996. *Interaction and grammar.* Cambridge: Cambridge University Press.
Ogden, R. 2009. *An introduction to English phonetics*. Edinburgh: Edinburgh University Press.
Ogden, R., and J. Local. 1994. Disentangling autosegments from prosodies: A note on the misrepresentation of a research tradition in phonology. *Journal of Linguistics* 30: 477–498.
Powell, M. 1996. *Presenting in English*. Hove: Language Teaching Publications.
Psathas, G. 1995. *Conversation analysis*. Thousand Oaks: Sage.
Richards, J.C. 2005. *Communicative language teaching today*. http://www.professorjackrichards.com/work.htm
Richards, K., and P. Seedhouse, eds. 2005. *Applying conversation analysis*. Basingstoke: Palgrave Macmillan.
Roth, W.-M., and K. Tobin. 2010. Solidarity and conflict: Aligned and misaligned prosody as a transactional resource in intra- and intercultural communication involving power differences. *Cultural Studies of Science Education* 5(4).
Sacks, H. 1992. *Lectures on conversation*. Oxford: Blackwell.
Savignon, S. 1991. Communicative language teaching: State of the art. *TESOL Quarterly* 25: 261–277.
Schegloff, E.A. 1998. Reflections on studying prosody in talk-in-interaction. *Language and Speech* 41: 235–263.
Schegloff, E., I. Koshik, S. Jacoby, and D. Olsher. 2002. Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics* 22: 3–31.

Seedhouse, P. 2005. Conversation analysis and language learning. *Language Teaching* 38(4): 165–187.
Seidlhofer, B. 2004. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics* 24: 209–239.
Seidlhofer, B., and C. Dalton-Puffer. 1995. Appropriate units in pronunciation teaching: Some programmatic pointers. *International Journal of Applied Linguistics* 5(1): 135–146.
Selting, M. 1995. *Prosodie im Gespräch. Aspekte einer interaktionalen Phonologie der Konversation*. Tübingen: Max Niemeyer.
Selting, M. 1996a. On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics* 6: 357–388.
Selting, M. 1996b. Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished' questions in repair initiation. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 231–270. Cambridge: Cambridge University Press.
Selting, M., and E. Couper-Kuhlen, eds. 2001. *Studies in interactional linguistics*. Amsterdam: Benjamins.
Selting, M., P. Auer, B. Barden, J.R. Bergmann, E. Couper-Kuhlen, S. Günthner, C. Meier, U. Quasthoff, P. Schobinski, and S. Uhmann. 1998. Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173: 91–122.
Sidnell, J., ed. 2009. *Conversation analysis: Comparative perspectives*. Cambridge: Cambridge University Press.
Skidmore, D. 2008. Once more with feeling: Utterance and social structure. *Text and Talk* 28(1): 79–96.
Spada, N. 2007. Communicative language teaching: Current status and future prospects. In *International handbook of English language teaching. Part 1*, ed. J. Cummins and C. Davison, 271–288. New York: Springer.
Szczepek Reed, B.B. 2004. Turn-final intonation in English. In *Sound patterns in interaction: Cross-linguistic studies from conversation*, ed. E. Couper-Kuhlen and C.E. Ford, 97–118. Amsterdam: Benjamins.
Szczepek Reed, B.B. 2006. *Prosodic orientation in English conversation*. Basingstoke: Palgrave McMillan.
Szczepek Reed, B.B. 2009a. Prosodic orientation: A practice for sequence organization in broadcast telephone openings. *Journal of Pragmatics* 41(6): 1223–1247.
Szczepek Reed, B.B. 2009b. FIRST or SECOND: Establishing sequential roles through prosody. In *Where prosody meets pragmatics*, ed. D. Barth-Weingarten, A. Wichman, and N. Dehé, 205–222. Emerald: Bingley.
Szczepek Reed, B.B. 2010a. *Analysing conversation: An introduction to prosody*. Basingstoke: Palgrave Macmillan.
Szczepek Reed, B.B. 2010b. Prosody and alignment: A sequential perspective. *Cultural Studies of Science Education* 5(4): 859–867.
Szczepek Reed, B.B. 2010c. Intonation phrases in natural conversation: A participants' category? In *Prosody in interaction*, ed. D. Barth-Weingarten, E. Reber, and M. Selting. Amsterdam: Benjamins.
Szczepek Reed, B.B. 2010d. Units of interaction: Tone units or turn constructional phrases? In *Conference Proceedings: Interface Discourse and Prosody*, ed. E. Delais-Roussarie. Paris: University of Chicago, 9–11 Sept 2009.
Szczepek Reed, B.B. forthcoming. Prosody, syntax and action formation: Intonation phrases as 'action components'. In: *Interaction and usage-based grammar theories*, ed. P. Bergmann and J. Brenning. Berlin: de Gruyter.
Tarplee, C. 1996. Working on young children's utterances: Prosodic aspects of repetition during picture labelling. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 406–435. Cambridge: Cambridge University Press.
ten Have, P. 2007. *Doing conversation analysis: A practical guide*, 2nd ed. London: Sage.
Tench, P. 1988. *The roles of intonation in English discourse*. Frankfurt: Peter Lang.

Thompson, S. 1995. Teaching intonation on questions. *ELT Journal* 49(3): 235–243.
van den Branden, K., M. Bygate, and J.M. Norris, eds. 2009. *Task-based language teaching: A reader.* Amsterdam: Benjamins.
Wagner, J. 1996. Foreign language acquisition through interaction: A critical review of research on conversational adjustments. *Journal of Pragmatics* 26: 215–235.
Walker, G. 2004. On some interactional and phonetic properties of increments to turns in talk-in-interaction. In *Sound patterns in interaction*, ed. E. Couper-Kuhlen and C.E. Ford, 147–170. Amsterdam: Benjamins.
Wells, J.C. 2006. *English intonation: An introduction*. Cambridge: Cambridge University Press.
Wells, B., and J. Corrin. 2004. Prosodic resources, turn-taking and overlap in children's talk-in-interaction. In *Sound patterns in interaction: Cross-linguistic studies from conversation*, ed. E. Couper-Kuhlen and C.E. Ford, 119–144. Amsterdam: Benjamins.
Wells, B., and S. Macfarlane. 1998. Prosody as an interactional resource: Turn-projection and overlap. *Language and Speech* 41: 265–294.
Wells, B., and S. Peppè. 1996. Ending up in Ulster: Prosody and turn-taking in English dialects. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 101–130. Cambridge: Cambridge University Press.
Widdowson, H. 1978. *Teaching language as communication*. Oxford: Oxford University Press.
Wong, J. 2002. "Applying" conversation analysis in applied linguistics: Evaluating dialogue in English as a second language textbooks. *International Review of Applied Linguistics* 40(1): 37–60.
Wong, J., and H.Z. Waring. 2010. *Conversation analysis and second language pedagogy: A guide for ESL/EFL teachers*. New York: Routledge.

Part III
Pedagogical Implications for English Language Teaching

Chapter 11
Same but Different: The Pragmatic Potential of Native *vs.* Non-native Teachers' Intonation in the EFL Classroom

Silvia Riesco-Bernier

11.1 (De)Constructing Meaning

Back in the early days of pragmatics, Austin (1962) associated the different speech acts with specific utterances (the performative verbs). However, Searle (1976) felt Austin's classification responded to a mere categorisation of English illocutionary verbs and understood that the basic semantic differences may have syntactical consequences (not only at verb choice level). Thus, Searle showed how the different basic illocutionary types are realised in the syntax of a natural language such as English. The existence of a wide range of linguistic realisations that enables the speaker to instantiate meaning(s) is related to one of the key notions in pragmatics, namely the "continuous making of linguistic choices" (Verschueren 1999:55, my inverted commas). The speakers, consciously or unconsciously, make choices which can be situated at any level of linguistic form (phonological, morphological, syntactic, lexical or semantic).

Interestingly enough, the exploration of the form-function relationship has recently allowed computational linguists to create computer models that consist of a speech tagger, a syntactic parser, a symbolic post-processor and a model based on surface linguistic structures, which altogether classify speech acts automatically, e.g. "The Auto-Tutor Programme" (cf. Graesser et al. 2001) and other programmes (Nagata and Morimoto 1994; Cohen and Shiverly 2003; Cohen and Ishihara 2004). What is more, some actually create language and perform speech acts, e.g., "Elephant 2000" (cf. McCarthy 1998).

However, a major issue within *Speech Act Theory* (hence, *SAT*) is the phenomenon of indirect speech acts. Bearing in mind that the illocutionary act or speech act is associated by convention with the form of the utterance in question, there is a literal

S. Riesco-Bernier (✉)
English Department, Universidad Autónoma de Madrid, Madrid, Spain
e-mail: sylvie.riesco@educa.madrid.org; s.riesco@uam.es

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_11,

force hypothesis (Gazdar 1981) whereby (i) explicit performatives have the force named by the performative verb in the matrix clause and (ii) the three major sentence-types in English, namely the imperative, interrogative and declarative, have the forces traditionally associated with them, i.e., ordering (or requesting), questioning and stating. However, when a sentence fails to have the force associated with (i) and (ii) above, the utterance has a literal force together with an inferred indirect force, which will be known as "indirect speech act" (cf. Searle 1975; Geis 1998; Levinson 1994). Such mapping between the linguistic surface structure and its subsequent meaning urges the linguist to consider not only the discourse-grammar interface but include a third component: phonology. Within the strata of language, phonology accounts for the division of information into workable units (tonality system), the placement of physical/acoustic prominence to the newest or most relevant information (tonicity system) and the intention or meaning conveyed by a particular pitch movement (tone system). Therefore, studies centred upon spoken discourse cannot but consider this layer as crucial in the communication of meaning.

While some criticisms question the truth value of some of the concepts posited by *SAT*, others are concerned about the nomenclature (Leech 1983; Levinson 1994; Verschueren 1999). In other words, *SAT* uses lexical labels to categorise verbal realities which "make fuzzy category distinctions, whereas the realities to which these categories apply are often scalar or indeterminate" (Leech 1983:225). Indeed, the lack of systematicity is reflected in a theory whose distinct categories are not exclusive since some utterances/acts could be hybrids (Verschueren 1999:24), which calls for a more flexible theory. When analysing speech acts, Levinson (1994:280) urges the reader to bear in mind the following disciplines: (i) the ethnography of speaking focused on cross-cultural study of language usage (cf. Bauman and Sherzer 1974) and (ii) language acquisition studies (cf. Bruner 1975; Dore 1975; Bates 1976; Snow 1979). Hence, the analysis of language in human communication should consider *SAT* together with "more complex multi-faceted pragmatic approaches" (Levinson 1994:278).

11.2 Intonation as a Meaning Maker Device

> "Whenever we describe a language, we are concerned with meaning, and all contrast in meaning can be stated either in grammar or in lexis. If we regard intonation in English as meaningful[…] we should seek to state the place which such choices occupy relative to the total set of formal patterns in the language" (Halliday 1967:11)

Intonation is not only a tool that encodes the structure and organisation of discourse but it also constitutes a meaning making resource: "intonation as a repertoire of conventions available to and interpretable by users of the language in just the same way as are the conventions of syntax" (Sinclair and Brazil 1982:94). Although this function has been acknowledged by many authors (Austin 1962; Searle 1969; Salaberri 1999), it has scarcely been studied in depth (Geluykens 1987). In this sense, intonation has not been examined as a system on its own (except in the case

of theoretical linguists) but as a constituent that is part of the system of language, relating thus intonation and meaning in some cases and/or intonation to syntax or grammar in others (Halliday 1967; Chafe 1987).

Although many systems of intonation co-exist in the literature, most take the following defining characteristics of the tones: the pitch direction (falling, rising), the complexity of the movement, the range of the movement (wide, narrow) and the height of the movement (high, low, mid) (Fox 2001:313). Palmer (1922) establishes four different tones: fall, high rise, rise-fall-rise and low rise; while O'Connor and Arnold (1961) offer a wider variety: low fall, high fall, rise fall, low rise, high rise, fall rise.

Within the Systemic Functional approach to language (hence, SFL), Halliday (1967, 1970, 1994) and Tench (1996) consider a five primary tone system based on the idea that constituency, which accounts for the organisation of language making its different units relate to each other by a part-whole relationship, also applies in speech. The rhythm and melody of speech account for the presentation of information in wider units that embed smaller units inside, contributing to the efficient communication of information. Further, three subsystems are at work within the intonation system when presenting information: (i) "tonality" is responsible for organising the information by dividing the text into blocks of message or units of information (i.e., tone group); (ii) "tonicity" organises the information within the tone groups by considering the two functions of information (given and new) and is mainly concerned with the focus of information (i.e., tonic); and (iii) "tone", responsible for the melody and pitch contours under which information is presented and thus assigns various meanings to distinct realisations.

11.3 Child Directed Speech: A Prosodic Discourse

From a suprasegmental perspective, Child Directed Speech, hereafter CDS, is produced by a higher and wider pitch range, intensity modulation, a slower tempo and exaggerated prosodic contours (cf. Ferguson 1964; Garnica 1977), which results in the infants' preferred discourse. Bryant and Barrett (2007) argue that infants usually prefer to listen to CDS speech over adult speech and even prefer CDS in a foreign language to adult speech in the language they are accustomed to hearing (Fernald and Morikawa 1993), a preference that can even start at the prenatal stage (cf. DeCasper et al. 1994; Altmann 1997).

The acoustic and prosodic characteristics of CDS embody the simplifying, clarifying and expressive/affective functions of language (Ferguson 1977). Whereas the first two functions are motivated by "the desire to be understood and, possibly, to teach" (Brown 1977:4) and are reflected in the segmental and acoustic changes (Cooper et al. 1985; Nooteboom and Kruyt 1987), the affective function intends to express "affection with the capturing of the addressee's attention" (Brown 1977:4) and is mostly related to the prosodic changes (Stern et al. 1983; Mochizuki-Sudo 1991).

In the late 1970s and early 1980s numerous studies have focused on the analysis of the expressive function of intonation to evidence a direct relationship between contours and communicative functions when addressing the child (cf. Dore 1974; Halliday 1975; Bates 1976; Sachs 1977; Fernald 1989). Furthermore, the discrimination among various prosodic contours to which different communicative values are assigned starts when the infant is 4 (Chang and Trehub 1977) or 6 months of age (Stern et al. 1983). In the last decades, cognitive linguistics and cognitive psychology have gathered around the notion "affective prosody" those studies that analyse the vocal expression and communication of emotions in child/infant directed speech (Banse and Scherer 1996; Johnstone and Scherer 2000; Thompson et al. 2004; Bryant and Barrett 2007).

More specifically, a dichotomy has been established between the falling tones, the melodic contours associated to referring, labelling and informing and the rising contours which gain the child's attention and engage them in interaction (cf. Sullivan and Horowitz 1983). Within this major categorisation, some have focused on the prosodic realisation of particular functions: clarification requests (Rodríguez and Schlangen 2004, Edlund et al. 2005) declaratives and questions (Frota 2002), prohibitives *vs.* approvals (Fernald 1992), among others.

However, while unmistakably distinct, CDS is also regarded as a bridge between the child and the adult system of communication. According to Cruttenden, "an intonation system with adjusted pitch range and pitch height is being used to introduce the child to some of the *meanings* of the adult intonation system" (Cruttenden 1994:145, my italics). What seems to shape intonation is not the learner's age but rather the intention/meaning the speaker wishes to convey: Trainor et al. (2000) found that while acoustic analyses showed few differences between the Infant Directed and Adult Directed samples, diversity exists across emotions. Furthermore, emotions and meanings shape intonation regardless the language: Fernald (1992) described how the pitch contours pattern similarly in relation to communicative intentions across several languages: whereas prohibitive utterances are often characterised by low F0, narrow F0 range and staccato-like bursts, approval vocalisations generally have high average F0, wide F0 range and a prominent F0 rise-fall contour. Likewise, Bryant and Barrett (2007) study how to recognise intentions in infant directed speech and even argue for universals across languages, which encourages the study of prosody as a crucial meaning-maker in intercultural contexts.

11.4 Analysis

11.4.1 Materials and Method

The present study is based on authentic contextualised spontaneous speech coming from the UAM-Learner English Spoken *Corpus*, a longitudinal *corpus* of the oral interaction in the EFL classroom in diverse teaching contexts in Madrid

Table 11.1 *Corpus* data

Corpus data	
Sample size in minutes	68
Sample size in functions	1,241
Sample size in tone units	1,776
School A	Bilingual school: English used to teach all subjects
	Teachers: 2 native speakers of English
School B	English is used half an hour daily
	Teachers: 2 non-native speakers of English
Speakers: Teachers	4 female teachers in their early 30s:
	2 native speakers of English: school A
	2 non-native speakers of English: school B
Learners	5 year old children
	There were 23–25 children in each classroom
Language	English is the foreign language to most children in both schools except in very few cases in the bilingual school (School A).

(Romero-Trillo and Llinares-García 2001; Llinares-García 2002; Riesco-Bernier and Romero-Trillo 2008a, b). The analysed *corpus* focuses on the first year of the compilation and consists of four recorded sessions- video-taped (*SONY Handycam Video* Hi8 XR), aided by tape-recorders placed at strategic points in the classroom. Table 11.1 above depicts the *corpus* data.

Firstly, the video-taped sessions were orthographically and prosodically transcribed after an auditory analysis. The discourse was then divided into *tone units*, defined as a unit of information with one intonation contour where there is just one "new" focus of information, the tonic (Halliday 1967, 1970; Riesco-Bernier and Romero-Trillo 2008a). Later, an ulterior visualisation of the raw wave form, the auto-pitch, the magnitude and raw pitch graphs (*Speech Analyzer Software* v.1.5) helped to identify the prosodic contours (Halliday 1970). Finally, each *communicative function*, defined as the minimal unit of meaning materialised in an utterance which occurs at a particular "move" in an exchange, was identified and a tag was assigned to both the meaning conveyed and the prosodic contour displayed. The quantification of the data could then ensue by using the *Wordsmith Tools Software v.3.0* (Scott 1998), which was ultimately followed by the statistical analyses carried out with the *SPSS Software v.10.0*.

11.4.2 Objectives and Hypotheses

The objective of this investigation is to explore the relationship between "meanings" and "the prosodic realisation" in EFL teacher talk from a theoretical and pedagogical perspective. First, this study offers an analysis of the prosodic repertoires in the

instantiation of communicative functions in EFL teacher talk. And second, an in-depth analysis of native and non-native teachers' discourse unveils the different exploitation of the phonological system so as to communicate in the classroom. The aforementioned objectives can be specified in the following hypotheses:

- Hypothesis 1: *A greater variety of choice will be found in native teachers' prosodic production*.
- Hypothesis 2: *There will not be a one-to-one correspondence "meaning"-"tone" relationship*.
- Hypothesis 3: *Native teachers will display a greater exploitation of the phonological system in the classroom*.

11.4.3 Analysis of the Data

11.4.3.1 Discourse-Semantic Analysis

"Speaking is something that might more appropriately be called an interact: it is an exchange" (Halliday 1994:68). Then, the act of speaking becomes an interactive process where both participants (speaker and listener/ writer and reader) are involved and where their roles depend on each other's, which results in a wide range of different types of "interactions" contingent on the specific context. Halliday acknowledges that the two main variables that come into play in the definition of the different interactional contexts and in the definition of the primary speech functions are the *speech role* and the *commodity exchanged* in the interaction.

The present study finds its roots in Halliday (1985), Hasan (1996) and Martin (1992), who tackled the analysis of language at the discourse-semantic stratum. It departs from the original four primary speech functions (give *vs.* demand information and give *vs.* demand goods and services) so as to develop it into a *Communicative Function Network System* (cf. Riesco-Bernier 2003): a system that subsumes 36 functions within 7 major categories and that are specific to EFL teacher talk.

As regards the unit of analysis at this stratum of language, this paper considers the "communicative function", inherited and shaped by works that constitute the origins of the study of meaning: *Speech Act Theory* (Austin 1962; Searle 1969) and *Classroom Discourse Analysis* (Sinclair and Coulthard 1975), in ESL (Long and Sato 1983; Ernst 1994) and EFL contexts (Salaberri 1999; Llinares-García 2002).

11.4.3.2 Prosodic Analysis

The study of meaning at the phonological stratum consisted of several analyses: first the data under study were analysed ***technically***, i.e., the data were digitised at a sampling rate of 22 KHz, quantised at 16 Bits and was high-pass filtered (*Speech Analyser Software v. 1.5*). Then followed an ***auditory*** analysis to identify the tone

groups,[1] tonic and tone of each unit although the present article mainly focuses on the analysis of tone.[2]

Tones can have simple contours, just consisting of one main pitch movement -rising or falling- or they can have complex pitch patterns (i.e., combining two pitch movements, having jumps, stepping movements…). Following a functional approach, seven primary tones can be recognised in English: five simple tones and two compound tones (cf. Halliday 1970). The difference between them lies on the pitch movement in the tonic segment. Generally speaking, the falling tone (tone 1) is mainly used for statements, the high rising (tone 2) is common in questions, the low rising (tone 3) in turn is generally used in polite commands or incomplete statements, the falling-rising (tone 4) usually indicates personal opinion, reservation or concessions and finally the rising falling (tone 5), which often embodies surprise as well as reproach and two complex tones (13 – falling + low rising and 53 – rising-falling + low rising).

The network system of intonation portrayed in Fig. 11.1 below summarises the prosodic repertoire and the communicative potential of the various contours that may be displayed by the speaker. Not only does choice exist at discourse level (choosing *what* to say) but also at the phonological stratum of the language (*how* to utter the words). Indeed, as inheritors of Firthian Linguistics, Systemic Functional Linguistics gives priority to the system (as the name suggests). Language is conceived as "networks of interlocking options" (Halliday 1994:xiv): the network system of meaning presents an inventory of ways in which meaning can be realised and analysed, and where there is an array of choices that will determine which meaning is being instantiated through language.

Regarding the tagging of the analyses, it should be said that each utterance is divided into tone units, presented with a code that displays the <tone> and the tonic syllable underlined together, and with a code acknowledging the communicative function instantiated. Finally, for illustration purposes, an example of the falling tone taken from the *corpus* is portrayed in Fig. 11.2 below, with the wave form, the raw and the auto pitch graphs.

11.5 Results

As the selection of the data followed qualitative criteria, both corpora under study (native *vs.* non-native) differ in the frequencies of the communicative functions produced. A simple contingency table displays the frequencies of the data in order to

[1] The "tone group" (Palmer 1922; Armstrong and Ward 1926; O'Connor and Arnold 1961; Brazil 1975), "rhythm unit" (Pike 1945), "tone unit" (Halliday 1967; Crystal 1969), "intonation unit" (Chafe 1987), "intonation group" (Cruttenden 1997) or "pitch sequence" (Brazil 1975) are many of the terms encountered in the literature when referring to the division of speech into workable operationalisation of speech.

[2] cf. Riesco-Bernier (2003) for a full account of the intonation system.

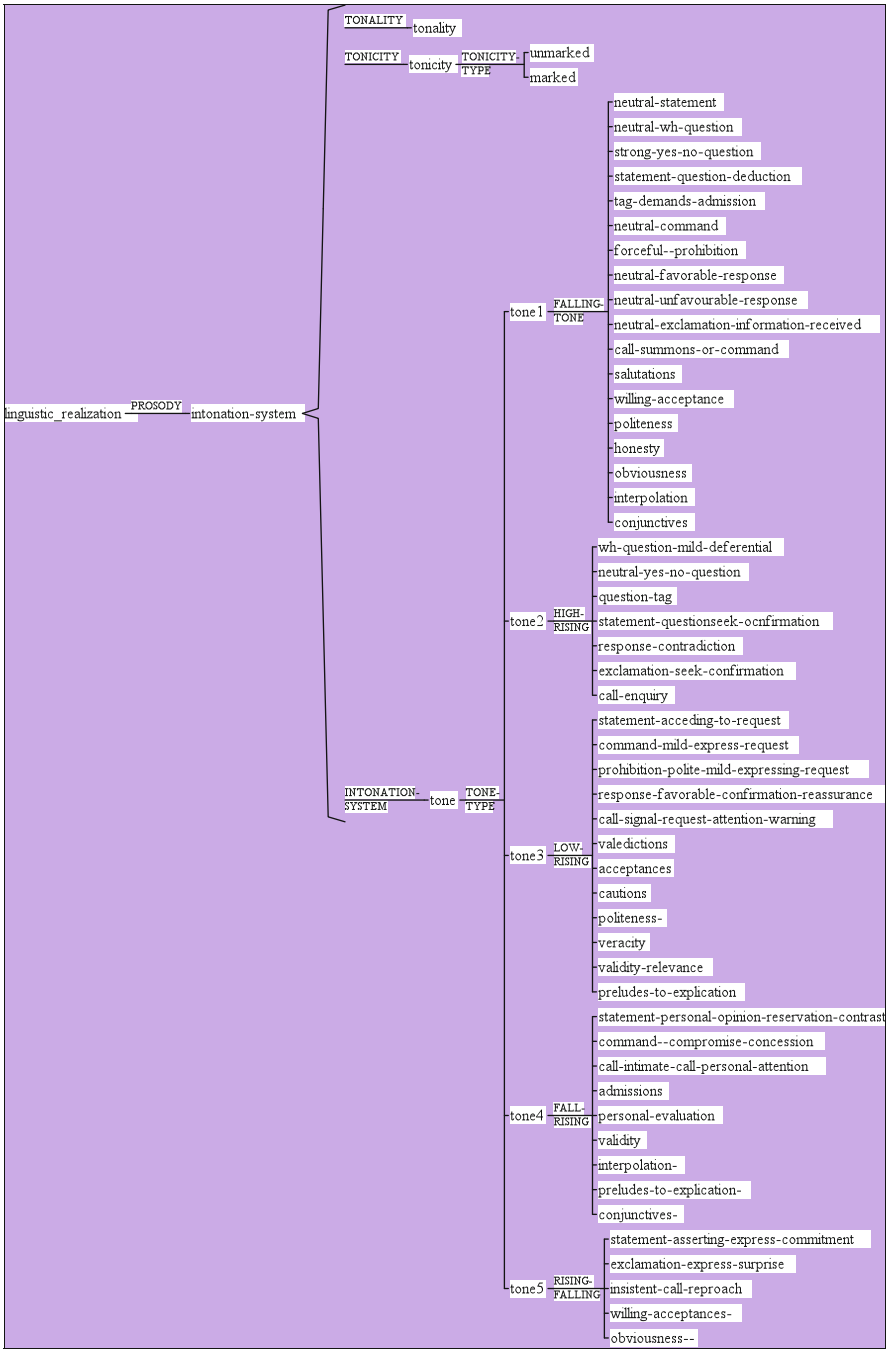


Fig. 11.1 Network system of the intonation system under SFL

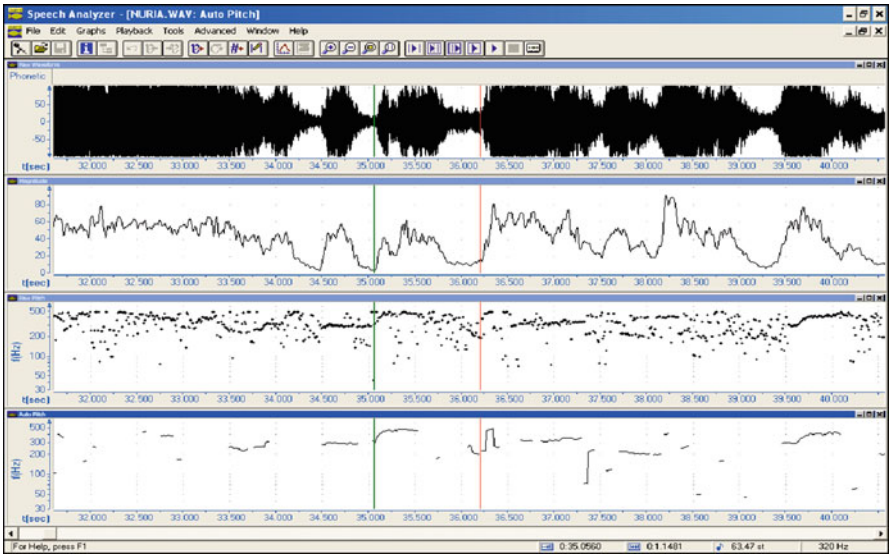


Fig.11.2 TONE 1: //1What colour is this?// (The prosodic transcription is provided, presenting the corresponding text to the space the cursors delimit. The double slashes indicate the tone units, the underlined syllable is the tonic elements and the tone appears at the beginning of the tone unit)

Table 11.2 Major speech function across teachers

	Language		
Primary speech functions	Native	Non-native	Total
Calls	136	150	286
Exclamations	14	9	23
Give information	204	223	427
Give goods and services	14	1	15
Demand information	83	109	192
Demand goods and services	143	147	290
Total	594	639	1,233

inform of the *corpus* size in relation to the seven major functions (Table 11.2 above). Hereafter, the results will be displayed in percentages and in relation to the specific communicative function.

11.5.1 Prosodic Realisation(s) of Communicative Functions in the EFL Classroom

The present section unveils the prosodic realisation(s) of the distinct communicative functions in teachers' discourse. These data have been summarised in contingency Table 11.3 below, which allows working with two qualitative

Table 11.3 Native vs. non-native teachers' prosodic repertoire (white vs. grey columns, respectively).

		Tone 1		Tone 2		Tone 3		Tone 4		Tone 5		Complex tone 13		Sequence tones 3+1
Communicative functions		NS	NNS	NS	NNS	NS	NNS	NS	NNS	NS	NNS	NS	NNS	NS
Calls	Nomination	–	4	7	6	2	3	5	–	1	–	–	–	–
	Selection	9	17	4	3	7	12	8	2	4	5	–	–	–
	Response to call	1	–	3	3	–	–	–	–	–	–	–	–	–
	Scolding	–	6	–	–	2	–	1	–	3	3	–	–	–
	Acknowledge	3	–	4	11	–	–	–	–	–	–	–	–	–
	Call attention	26	14	1	1	7	3	–	–	2	–	–	–	–
	Acknowledge response	24	39	–	–	8	9	–	–	1	4	–	–	–
	Challenge audience	2	1	–	4	–	–	1	–	–	–	–	–	–
Exclamation		6	4	3	2	2	–	–	–	3	3	–	–	–
Give information	Give personal info	2	4	–	–	–	1	–	–	1	–	–	–	–
	Give instruction	17	18	–	–	3	3	–	–	–	–	–	2	–
	Agree	4	10	–	1	–	3	–	–	–	–	–	–	–
	Positive feedback	20	45	–	–	7	3	3	–	13	2	–	–	–
	Disagree	1	1	–	–	1	–	1	–	1	–	–	–	–
	Negative feedback	12	4	4	7	–	1	1	2	1	–	–	–	–
	Informative	7	5	1	–	–	–	1	–	–	–	–	–	–
	Explain	13	14	–	–	2	–	2	–	–	–	1	–	–
	Give answer	9	13	–	–	4	–	–	–	–	1	–	–	–
	Restating feedback	8	17	–	–	1	1	–	–	–	–	–	–	–
	Corrective feedback	8	10	1	2	1	3	2	–	–	1	–	–	–
	Enhancing feedback	2	11	–	14	–	1	–	–	–	–	–	–	–
	Guiding feedback	39	12	7	6	3	5	–	–	–	–	–	–	–

Give goods & serv.	Offer	8	–	–	–	3	–	–	–	–	–	–	–	–
	Threat	2	1	–	–	–	–	–	–	–	–	–	–	1
Demand information	Confirmation request	–	1	2	7	–	–	–	–	–	2	–	–	–
	Clarification request	1	1	1	2	–	–	–	–	–	–	–	–	–
	Referential question	22	17	11	4	–	1	–	–	–	–	–	–	1
	Check	4	–	7	1	–	–	–	–	–	–	–	–	–
	Comprehension check	–	–	11	17	–	–	–	–	–	–	–	–	–
	Display question	19	41	4	14	–	1	–	–	–	–	–	–	–
Demand goods and services	Command	82	86	8	3	14	11	2	–	–	1	9	10	–
	Allow	3	–	–	–	–	–	–	–	–	–	–	–	–
	Prompt	5	23	–	1	4	–	–	–	–	–	–	–	–
	Prohibition	6	10	1	2	–	–	–	–	–	–	–	–	–
	Suggest	6	–	–	–	–	–	–	–	–	–	–	–	–
	Warning	2	–	–	–	1	–	–	–	–	–	–	–	–
	TOTAL	**373**	**429**	**80**	**111**	**72**	**61**	**27**	**4**	**30**	**22**	**10**	**12**	**2**

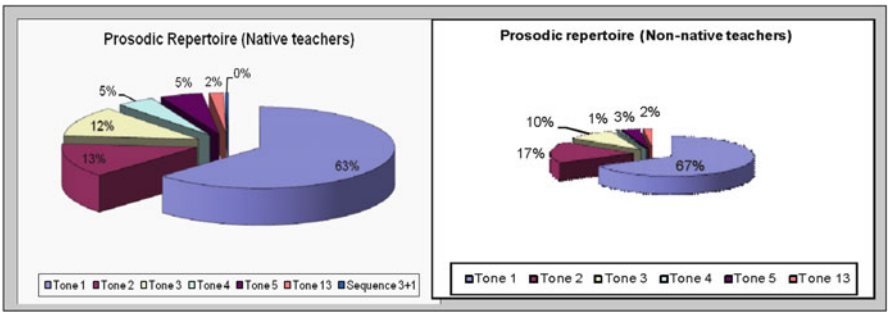


Fig. 11.3 Percentages of the prosodic repertoire in teachers' discourse

variables ("communicative function" and "tone") and presents the simultaneous distribution for the variables "tone" per "function" per "speaker" (native *vs.* non-native teachers).

For comparison purposes, Fig. 11.3 above portrays the use of the distinct prosodic contours regardless the communicative function instantiated in both groups. Percentages are used to appreciate the prosodic repertoire that teachers use in the classroom.

The data displayed above validate the first hypothesis posited in this paper: "a *greater variation in choice will be found in native teachers' prosodic production"*. We shall call "variation in choice" the degree to which the speaker (native *vs.* non-native teacher) displays different phonological structures to convey the same communicative function. First, the number of prosodic contours realised in both corpora differs: while native teachers display seven tones, non-native teachers use six. Note that the sequence of the tones 3 + 1 is scarcely used by native speakers in the instantiation of threats or referential questions but is non-existent in the non-native *corpus*. Further, Fig. 11.3 shows that a wider repertoire in the case of native teachers implies fewer instances produced within each tone: native teachers seem to use tones 1 (high fall) and 2 (high rise) less frequently than their non-native counterparts whereas their use of tones 3 (mild rise), 4 (fall-rise) and 5 (rise-fall) exceeds the non-native's production.

Second, Table 11.3 reveals that the tone system is further exploited by native teachers as in 31% of the communicative functions, native teachers used a wider repertoire of tones when realising *scolding calls, calls to acknowledge, calls of attention, exclamations, positive feedback, disagree, give information, explain, threat* or *offer*. Non-native teachers only exploited the tone system further in 13% of the functions (i.e., *give instruction, agree, give encouraging feedback, confirmation request and display questions*).[3]

[3] Both groups displayed the same variety of prosodic choice in the remaining 56% of the cases.

11.5.2 The Pragmatic Potential of Intonation

As a cross-stratal study of meaning (from discourse-semantics to phonology), this paper examines in this section the cross-stratal interaction of "prosodic realisations" and "communicative functions", so as to test (i) if there is a dependency relationship between both variables and (ii) if such dependency means a bi-uniqueness (a one-to-one relationship) and thus evaluate the pragmatic potential of intonation in teachers' discourse.

11.5.2.1 The "Tone"-"Communicative Function" Relationship

The exploration of the dependency between the two variables can first be appreciated through an impressionistic reading of Table 11.3 above: there is an unequal distribution of the frequencies, since they tend to concentrate on some phonological realisations. Each row (regulatory functions) has one cell where the percentage of frequency with the column, (i.e., prosodic realisation = tone) is higher than in the rest. Should the two variables (function and tone) be independent, the distribution would be equitable in their distribution.

Second, in order to consider whether both variables are statistically related, this study considered the *Likelihood Ratio Chi-Square* and the *Pearson Chi-Square* coefficients, which will examine the null hypothesis that the two variables are independent. Since the *p* value obtained in this case is $p = .000$ both for native and non-native teachers, the null hypothesis is rejected at α level of significance (=0.05). In other words, there is a statistically significant dependency between the "communicative function" and the "tone" in the native and the non-native groups of teachers. This dependency between the two variables is confirmed by the *p* values attached to the *Cramer's V* coefficient, ($p = 0.000$, for the native and the non-native teachers) which account for the statistical significance of the association. Furthermore, the values attached to the *V* coefficient manifest that the degree of this relationship is not very high but still considerable in both groups (0.415 for the native teachers and slightly lower for the non-native group, namely, 0.413) as *Cramer's V* ranges from 0 when no dependency among variables exists, to 1 when they are perfectly related.

Another analysis further checked the dependency relationship through a measure of association for nominal-level variables, i.e., the *Uncertainty Coefficient*, which accounts for the direction of the dependency of two variables. This measure can be interpreted as the proportion in which the uncertainty in predicting the values of the dependent variable (in this case, the regulatory function) is reduced when considering the information from the independent variable (in this case, the prosodic contour). This coefficient ranges from 0 to 1, which indicates a complete reduction of error in predicting the dependent variable.

The values of the *Uncertainty Coefficient* were =0.124 and =0.117 for the native and non-native group, respectively. In other words, information about the

Table 11.4 Communicative potential of intonation

	Communicative functions native teachers	Communicative functions non-native teachers
Tone 1	32	28
Tone 2	18	21
Tone 3	18	16
Tone 4	11	2
Tone 5	10	9
Tone 13	3	2
Tones 3 + 1	2	0

tone would help in reducing the error in the prediction of the function in 12.4% in the case of the native group and even less (11.7%) in the non-native group. On the contrary, if the direction of the association is inverted, (cf. "T dependent values"), the values are much higher in both groups (0.317 and 0.349 for the native and the non-native group), which indicates that knowing the communicative function helps in discriminating the tone (and not the other way around), mostly in the non-native group.

Consequently, despite the low coefficients attached to the strength and direction of their association, it is possible to claim that there is a statistically significant dependency relationship between the variables "tone" and "communicative function".

11.5.2.2 Dependency *vs*. Bi-uniqueness: The Communicative Potential of Prosody

Had the values obtained (*Crammer's V and Uncertainty coefficient*) been =1, the perfect correlation between "tone" and "communicative function" would result in bi-uniqueness, the relationship whereby a certain prosodic contour conveys one and only one communicative function. However, the results obtained in this study reveal that both variables are statistically related to each other but go beyond one-to-one correspondence: contingency Table 11.3 reveals that only 7.9% of the functions were displayed by a unique phonological realisation. Furthermore, Table 11.4 above summarises the number of functions instantiated by each prosodic contour in both groups of teachers.

Additionally, Figs. 11.4 and 11.5 below present a multiple bar graph whose aim is to compare among categories the percentage each tone brings to the total. In both graphs, each bar stands for a tone and illustrates with different colours the different communicative functions conveyed with their percentages in both groups of teachers.

The findings displayed validate the second hypothesis of the present study, i.e., "*there is not a one-to-one correspondence between the communicative functions and their phonological realisation*" for most functions (92,1%). This lack of neat fit

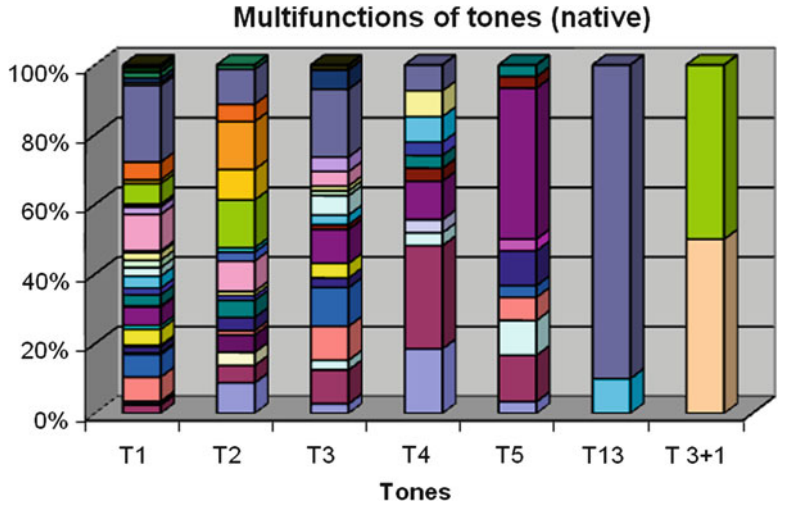


Fig. 11.4 Bar graph "tones" and "functions conveyed" in native teachers

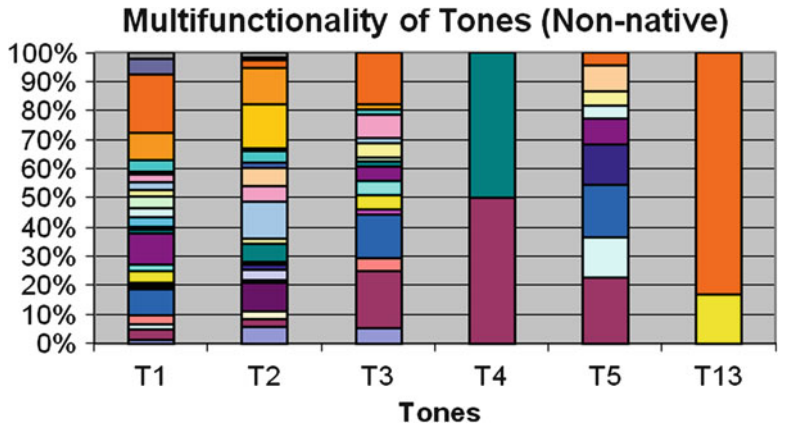


Fig. 11.5 Bar graph "tones" and "functions conveyed" in non-native teachers

highlights the potential of language and evokes the idea of "choice", whereby the speaker tends to use one or other structure to convey some meaning.

11.5.2.3 A Different Exploitation of Prosody

The data presented so far have already hinted at similarities and differences in native *vs*. non-native teachers' exploitation of prosody. The present section summarises in a comparative chart the findings obtained so far and leads us to test hypothesis 3,

i.e., "*The intonation system is exploited differently by native and non-native teachers*", and validate it with reservations as similarities co-exist with differences in the teachers' prosodic realisation(s) of meanings regardless their L1 (Table 11.5).

11.6 Discussion of Findings and Pedagogical Implications

The tone system displays a whole array of choices leading to the production of a specific communicative function, where the speaker stands as an active participant whose choices contribute to a personal communicative style. This section discusses the similarities and differences between native and non-native teachers' production and provides pedagogical implications.

11.6.1 Child-Directed Speech: A Discourse with Similar Traits Across Speakers

11.6.1.1 The "Phonological Realisation-Communicative Function" Relationship

The lack of neat fit found in Sect. 11.5.2 between those two variables highlights the potential of language, which can display a limited set of structures in an uncountable number of ways to shape meaning. First, while no one-to-one correspondence between the surface and meaning existed, it was also the case that in most functions in both corpora (native and non-native), one of the patterns at the phonological plane predominated over the rest, those being the ones acknowledged in the literature as the "unmarked" structures.

The *tendency* for a specific linguistic pattern to instantiate a particular communicative function can be interpreted in the light of *Prototype Theory* (Rosch 1977, 1978; Radden 1992). The concept of prototype accounts for an explanation to the problem of categorisation and category membership where natural categories are usually defined in terms of prototypes which combine the most representative attributes of a category, the prototype being the best, most salient (most frequent) amongst the members of the category and standing as the cognitive reference point. Further, the prototype is related to the surrounding members (less prototypical) in the extent to which those share traits and features. Moving to our results, it can be argued that there is indeed a prototypical phonological realisation of each communicative function, which corresponds to the most frequent displayed pattern and that the resemblance principle also applies. Actually, when several prosodic realisations were produced by the speaker in order to instantiate a "nomination", for example, the prototypical contour being the rising tone (tone 2), the second most frequent contour was the fall-rise (tone 4), and then the slight rise (tone 3), the final one being

Table 11.5 Comparison of native vs. non-native teachers' use of intonation

		Native teachers	Non-native teachers
Similarities	Although no bi-uniqueness exists in 92.1% of the functions, there is a tendency for a prosodic contour to predominate over the rest in each communicative function.		
	The chosen prosodic contour that instantiates a particular communicative function coincides in both groups in 81.5% of the cases.		
	It is in the case of Comprehension checks, Suggestions and Allowances that a single prosodic contour is displayed.		
	Both exploit the first three prosodic contours (high fall, high rise and mild rise) to a wider extent.		
Differences	Prosodic repertoire	They display 7 prosodic contours. In 31% of the functions, they display wider variety of tones.	They display 6 prosodic contours. In 13% of the functions, they display wider variety of tones.
	Exploitation of the *tone* system: choices	Tones 13, sequence tones 3+1, fall rise (tone 4) and rise-fall (tone 5): mainly used by native teachers.	Sequence tones 3+1 inexistent. Tone 2 overused.
	Potential of "*Tone*": Multifunctionality of tones	Their tones embody a wider variety of communicative functions (see Table 11.3).	Their tones embody a smaller variety of communicative functions except for the rising tone (tone 2).
		Strength of association of the variables function-tone is higher (Crammer's $V = .415$ vs. non-native's .413)	Higher value of reducing the error in the prediction of the tone: (*Uncertainty coefficient* = .349 vs. native's .317).

the rise-falling contour (tone 5). Analysing the four different prosodic realisations, ordered in decreasing order of frequency, they involve a rising pitch, to greater or lesser extent, with the exception of the latest (tone 5, which involves a rise but ends on a falling contour). Not surprisingly, the rise-fall occupies the final position. In other words, the different frequencies obtained in the structures displayed for the same communicative function in the data are not arbitrary but respond to an order established by the degree of likeness to the prototype, i.e., the closer to the prototype, the more frequent its use.

A word of caution must, however, be provided since the relationship "clause-type" and "tone" is to be considered. As the direction in communication goes from discourse-semantics to lexicogrammar and in turn to phonology, some prosodic realisations are indeed determined by the grammatical surface embodying the function. The case of "display questions" or "referential questions" may well illustrate this interaction: while those are prototypically instantiated by a rising tone, if their lexicogrammatical realisation is a *wh-question*, a falling tone will prevail. This does not constitute the exception to the "prototype" rule, it once again evidences that the three strata of language (discourse-semantics, lexicogrammar and phonology) interact together.

Second, the results of the phonological analysis of the communicative functions show that a statistical relationship exists between the functions and their tones although the coefficient associated with reducing the error of prediction is low. Riesco-Bernier (2003) found that tonicity and tonality are also statistically related to the communicative functions but their coefficients of association are much lower. It could be argued that these findings respond to the different roles each sub-system in intonation plays in communication. It is widely accepted in the literature that intonation supplies two major functions in communication: an interpersonal and a textual function. The *tone* is related to the interpersonal metafunction of language. Previous research highlights the communicative value of intonation assigns the neutral unmarked patterns of intonation for different communicative functions (O'Connor and Arnold 1961; Halliday 1970; Tench 1996) and explores the relevance of prosody as a communicative cue in child-directed speech (Dore 1974; Halliday 1975; Garnica 1977). *Tonality* and *tonicity* instead, contribute to the textual metafunction of language by structuring discourse (Goodenough-Trepagnier and Smith 1977; Swerts and Geluykens 1993; Sluijter and Terken 1993) and by signalling the most relevant information (Nooteboom and Kruyt 1987; Eefting 1991; Ladd 1996), both oriented to enhance speech perception and comprehension of information. This would then account for "tone" to be the phonological feature with a highest degree of association with the communicative function in the present findings.

11.6.1.2 A Common Aim: A Simple and Efficient Discourse

The qualitative analysis of the data and the ulterior results presented evidence that the *degree* of exploitation (selection of *choices*) responded to the easiness/difficulty intrinsic to the different possible choices within the *Intonation* system. In other

words, a relation can be found between the extent to which teachers exploited the tone system and the complexity to display certain choices. "Complexity" should here be understood as the difficulty the non-native teachers may experience in the production of certain patterns and also as the intricacy certain patterns convey to a young audience, for which English is the foreign language too.

If the phonological realisations of the two groups of speakers are closely observed, it can be claimed that all teachers avoided complexity in both schools. The analysis of the *tone* system reveals that those contours involving a single pitch movement, i.e., the fall (tone 1), the high rise (tone 2) and the slight rise (tone 3), are the most frequent in the two groups of speakers (cf. Sect. 11.5.1). And, among them, the falling tone was by far the favourite prosodic contour (62% in the native teachers and 69% in the non-native teachers), followed by the rising contour in both corpora. Should we indeed add up these two prosodic realisations, we would account for more than three quarters of the data, which hints at the *dichotomy* of the tone system in our teachers. First, these findings seem to echo the basic division between falling and rising pitch contours acknowledged in the literature (Armstrong and Ward 1926; Brazil 1975) and that emerged from the need to systematise the analysis of intonation:

> the basic division between falling and rising tunes arose because obviously this was the simplest type of formal division for analysts to make, but it also arose because the relationship between intonation and grammar was taken as the central fact to be described (Cruttenden 1981:78).

Moreover, analysts that have attempted an intonational description in more attitudinal terms (cf. O'Connor and Arnold 1961; Halliday 1967) have developed more complex patterns (ten and five simple tones, respectively) that nonetheless centre upon the two major prosodic contours. It then follows that the basic oppositions assigned to the textual/discursive and attitudinal/communicative functions assigned to intonation also present a dychotomic nature (e.g. "referring" *vs*. "proclaiming" tones; "certain" *vs*. "uncertain" tones; "reinforcing" *vs*. "limiting" tones, etc…). This might well account for a clear preference of use of the falling and rising contours (the more general and frequent in the system) by the teachers in our data.

Second, the register that has been here analysed is child-directed speech, which again accounts for simplicity in the linguistic structures displayed by the adult (the teacher). Halliday (1975) noted that two main prosodic contours are identified by the child, which s/he already displays at the first stage of his/her linguistic development: protolanguage. The falling contour is associated with the mathetic function of language (language used to tell and learn about the world) whereas the rising contour is related to the pragmatic function of language (language used to interact, to gain the other's attention…). Further research corroborates this tendency in the adult speaker interacting with infants and children, and associate "labelling", "referring" and "stating" functions to falling tones, *vs*. "demanding", "requesting", "calling attention" assigned to rising tones (Sachs 1977; Stern et al. 1983; Sullivan and Horowitz 1983). Actually, these patterns are confirmed by the results of our research, where the falling tone predominates in the *giving information* functions

(personal information, instructions, agreeing, any feedback, informatives, explicatives, answers) while the rising tones predominate in *interactional* oriented functions, namely the *demanding information* (confirmation requests, clarification requests, checks) and most *calls* demanding a verbal response – calling attention-(nomination, response to call). As for the demanding goods and services, it is interesting to note that the falling and rising contours coexist. This might be explained in the light of (i) the lexicogrammatical-phonological strata interaction (imperatives carrying a falling tone, interrogatives mainly embodied in rising tones) and of (ii) child-directed speech. Indeed, Garnica (1977) demonstrated that the use of terminal rises in imperatives was more common in 2-year old children than in 5-year old children. We could then argue that the combination of the falling and rising contours could be the evidence of a transition stage in the teacher's speech (from child-directed to adult-directed).

11.6.2 Native and Non-native Child-Directed Varieties

11.6.2.1 The Exploitation of Prosody

While the findings reveal that there is an identical *dependency* relationship between the communicative function and its phonological realisations in both groups, the values assigned to the coefficients of the *strength* of the associations and the reduction in error when predicting the communicative function (*PRE*) portray differences within and across groups. Riesco-Bernier (2003) showed that the same native and non-native teachers analysed in the present paper displayed a stronger degree of association in "communicative function"-"lexicogrammatical features". The present findings confirm that children at the age of five better comprehend the lexicogrammatical than the prosodic cues in meaning. Nonetheless, the communicative value[4] assigned to phonological features differed across groups. Whereas the degree of association between the communicative function and the lexicogrammatical features and the corresponding *PRE* were stronger in the non-native group (school B), the strength of association found in this paper between the communicative function and the *phonological* features and the corresponding *PRE* were higher in the native group (school A).

A different degree of exposition and immersion (in)to English by the children of the different groups could stand as a possible reason accounting for these results. While the two classes taught by the native speakers had a full-time immersion into English, the two classes taught by the non-native teachers were only addressed in English half an hour daily. Considering the role of immersion (cf. Cenoz and Perales 2000) and of input in the ulterior child's comprehension and acquisition of

[4] We shall refer to "communicative value" in this section as the degree of association between the interaction "communicative function" and "prosodic realisation".

linguistic skills (cf. Barnes et al. 1983; Kloth et al. 1998), it could be argued that those teachers interacting with children with a less proficient level of English (the non-native teachers) displayed easier and more explicit structures (lexicogrammatical cues) in conveying a particular communicative function rather than making use of intonation.

Indeed, despite being one of the first acquired linguistic skills in the mother tongue, intonation is one of the latest in the case of the second or foreign language. Although the children under study are very young, which can leads us to hypothesise that their acquisition of English intonation will be faster than in other second or foreign language contexts (e.g., 10–12 years old children), time of exposition constitutes a crucial factor enhancing their linguistic abilities. This can allow us to presuppose that the children with native teachers (school A) indeed have a greater proficiency of the prosodic system (in perception terms), which will account for those teachers displaying a stronger association of tones with the communicative function conveyed in the native group.

11.6.2.2 Variation in the Display of Choices

As shown in Sects. 11.5.1 and 11.5.2 above, no one-to-one correspondence exists between the "communicative function" and the "phonological realisation" in 92.5% of the data, which highlights the potential of intonation. Among the whole range of possible prosodic realisations, the choices the speaker makes reveal the main differences in the use and exploitation of the tone system by native *vs*. non-native teachers in the EFL classroom.

As mentioned above, "variation in choice" refers to the degree to which the speaker (native vs. non-native teacher) displays different phonological structures to convey the same communicative function. The results in Sect. 11.5.1 reveal that native and non-native teachers display the same number of phonological patterns in 56% of the communicative functions. However, in the cases where one group displays more structures than the other group, it is the native group that overrates the non-native group at the phonological stratum (31% of the data). This may well explain why the degree of reducing the error of predicting the tone once the function is known is lower in the native group: the more variety is found in the prosodic realisation of functions the less predictable the native teachers' discourse will be.

It could here be claimed that the potential of language depends on the lack of bi-uniqueness "meaning"-"form", but also on the speaker's (teacher) and the listener's (child) knowledge and mastery of the language. In other words, a neat fit between form and meaning would imply there is linguistic poverty in the system of communication (repeated structures for a same communicative function) but would guarantee an easy, practical and systematic teaching and learning of structures to communicate. On the contrary, the existing variation in choice found in our data, proper to any natural language, implies that both the speaker and the listener must know the different possible manners the speaker may use to communicate. Consequently, it could be argued that native teachers display a wider amount of both

phonological patterns because they are *native* speakers of English, thus more proficient in English and because of their audience. Indeed, the children taught in school A by the native teachers are more trained to recognise and understand a wider variety of patterns indicating the same communicative function.

11.6.2.3 Pedagogical Implications

The results above indicate that the degree of association between the "communicative function"-"tone" within the non-native teachers, the values assigned to the Cramer's V coefficients were lower than in the native speakers. Likewise, their exploitation (variation in choices) of the intonation system could be interpreted as being "poorer" than the natives'.

Non-native teachers display fewer prosodic patterns than the native teachers in order to convey the same communicative function (Table 11.4). Furthermore, as shown in Fig. 11.3, the tendency to display the falling (tone 1) and rising (tone 2) contours is much stronger than that found in the native *corpus*, which might be due to the avoidance of displaying complex tones (the fall-rise (tone 4) was only used in 0.6% of their instances *vs*. the 4.5% of the native teachers'). Indeed, such a limited use of the fall-rise (tone 4) accounts for a wider use of the rising contour (tone 2) in the non-native teachers. Likewise, although the rise-fall contour was sometimes displayed in the non-native *corpus*, its use was inferior to the one in the native teachers, which might well account for a "substitution" of the rise-fall (tone 5) by the display of the falling pitch (tone 1).

However, despite the underuse of some prosodic contours within the tone system, two factors argue in favour of the non-native teachers' exploitation of the phonological system. In other words, its different exploitation of the intonation system does not necessarily imply being a poorer one. First, the tendency to avoid complex tones and its "substitution" by simple contours might be interpreted as a strategy in communication. Indeed, Fernald and Kuhl (1987:280) argue that psychoacoustic research with adults suggests that the relatively simple pitch contours typical of motherese, considered as auditory patterns, may be processed and remembered more efficiently than the more complex and variable pitch contours. It would then seem that non-native teachers (school B) overuse the fall and rising tones (tones 1 and 2), since they would abstract the falling-rising opposition and would produce the more general (and easiest) contours. In fact, Fig. 11.5 reveals that whereas some tones are rarely communicative, tones 1 and 2 have a much wider degree of multifunctionality, i.e., the extent to which a tone embodies distinct communicative functions. This tendency and the overuse of the falling pitch was also acknowledged in a non-native *corpus* of Spanish university students preparing to be EFL teachers (Ramírez Verdugo 2003).

And second, it must be argued that the *tonic pitch height* constitutes a crucial cue in the non-native teachers' intonation system. The same data were analysed in previous research (Riesco-Bernier and Romero-Trillo 2008a, b) and revealed that the non-native teachers' pitch height was (i) statistically significantly higher than

the non-native teachers' and that (ii) its use was intended to fill the gap the tones created in their speech. In other words, the present analysis of the data and the statistical analysis of the present paper evidence that pitch movement may well be replaced by pitch height in the non-native speakers' talk. Indeed, the rare display of complex tones (fall-rise, tone 4, and rise-fall, tone 5) evidences the difficulty in producing pitch movement changes, which is further confirmed by the spectrogram analysis of the very basic falling (tone 1) and rising (tone 2) contours, where the pitch descent or ascent are by far simpler and sometimes hardly perceptible in contrast to the noticeable swiftness of the native's contours. It would then seem that tonic height intends to disguise the lack of mastery of the pitch movement.[5]

The so-called limited exploitation of the phonological system by the non-native teachers in relation to the native teachers' might respond to two major factors (i) either a lack of knowledge of the system of English, a finding that was also found in a study on tone choice and its discursive effects in international teaching assistants in the United States (Pickering 2001); or (ii) underline once again the differences between the children exposed to a greater or lesser input, which then shapes the discourse of the teacher (adapted to the child's comprehension). Although it is here felt that the second reason might well be the main factor, this has not been empirically studied in this research as yet, which then leads us to consider the first cause.

What the findings call for is the explicit teaching of the phonological systems in order for the teacher to be competent enough to enhance his/her linguistic abilities to communicate and in turn, teach learners to do so. Intonation has been consigned to oblivion in English Language Teaching (Morgan 1997; Levis 1999; Pickering 2001), which has led EFL learners to infer rather than learn the intonation patterns and their uses. Among the different pedagogical implications, raise-awareness techniques could well be the best option (Tyler 1992) in that they would allow the non-native teacher to first learn to discriminate among the different prosodic contours (speech perception) both within and across speakers (native and non-native), which will only then allow him/her to produce various patterns (speech production). The teachers of future EFL teachers should also encourage practice both in class and at home (self-study), which is nowadays possible by doing online exercises available on web (cf. references below).

11.7 Conclusion

There are three major conclusions that we may draw from the results of our analyses on the prosodic realisation of communicative functions in EFL teacher talk: first, there is a statistically significant dependency between the communicative functions instantiated in the classroom and their prosodic realisation. Second and fortunately,

[5] Cf. Riesco-Bernier (2003) for a comprehensive analysis of the pitch height in relation to the communicative functions.

such a relationship is not bi-unique: i.e., although tendencies exist in the production of certain contours to instantiate specific functions, there is not a one-to-one relationship (function-tone), which highlights the communicative potential of the tone system. Third, despite some similarities existing in the analysed subjects' discourse (teacher talk), both the variety of choices displayed and the complexity of the tones produced confirm that native and non-native teachers actually produce two varieties of Child Directed Speech.

What is left to investigate in the near future is whether the native teachers' discourse still stands as the richer and more elaborate standard to imitate and thus their prosodic repertoire is to be learnt and used by non-native speakers or if, on the contrary, non-native teachers are efficient in communication as their audience understands the various meanings conveyed.

References

Altmann, G.T.M. 1997. *The ascent of Babel*. Oxford: Oxford University Press.

Armstrong, L.C., and I.C. Ward. 1926. *A handbook of English intonation*. Cambridge: Heffer.

Austin, J.L. 1962. *How to do things with words*. Oxford: Oxford University Press.

Banse, R., and K.R. Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70: 614–636.

Barnes, S., M. Gutfreund, D. Satterly, and G. Wells. 1983. Characteristics of adult speech which predict children's language development. *Journal of Child Language* 10: 65–84.

Bates, E. 1976. *Language and context: The acquisition of pragmatics*. New York: Academic.

Bauman, R. and J. Sherzer, eds. 1974. *Explorations in the ethnography of speaking*. Cambridge: Cambridge University Press.

Brazil, D. 1975. *Discourse intonation*. Birmingham: University of Birmingham.

Brown, R. 1977. Introduction. In *Talking to children, language input and acquisition*, ed. C. Snow and C. Ferguson. Cambridge: Cambridge University Press.

Bruner, J.S. 1975. The ontogenesis of speech acts. *Journal of Child Language* 2: 1–21.

Bryant, G., and C. Barrett. 2007. Recognizing intentions in infant-directed speech. *Psychological Science* 18: 746–751.

Cenoz, J., and L. Perales. 2000. Las variables contextuales y el efecto de la instrucción en la adquisición de segundas lenguas. In *Segundas Lenguas. Adquisición en el Aula*, ed. C. Muñoz, 109–127. Barcelona: Ariel Lingüística.

Chafe, W. 1987. Cognitive constraints on information flow. In *Coherence and grounding in discourse*, ed. R.S. Tomlin, 21–51. Amsterdam: John Benjamins.

Chang, H.W., and S.E. Trehub. 1977. Auditory processing of relational information by young infants. *Journal of Experimental Child Psychology* 24: 324–331.

Cohen, A.D., and N. Ishihara. 2004. *A web-based approach to strategic learning of speech acts research report*. Minneapolis: University of Minnesota, The Center for Advanced Research on Language Acquisition.

Cohen, A.D., and R. Shiverly. 2003. Measuring speech acts with multiple rejoinder DCTs. *Language Testing Update* 32: 39–42.

Cooper, W.E., S.J. Eady, and P.R. Mueller. 1985. Acoustical aspects of contrastive stress in question-answer contexts. *The Journal of the Acoustical Society of America* 77(6): 2142–2156.

Cruttenden, A. 1981. Falls and rises: Meanings and universals. *Journal of Linguistics* 17: 77–91.

Cruttenden, A. 1994. Phonetic and prosodic aspects of baby talk. In *Input and interaction in language acquisition*, ed. C. Gallaway and B.J. Richards, 135–153. Cambridge: Cambridge University Press.

Cruttenden, A. 1997. *Intonation*. Cambridge: Cambridge University Press.
Crystal, D. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
DeCasper, A.J., J.P. Lecanuet, M.C. Busnel, C. Granier-Deferre, and R. Maugeais. 1994. Fetal reactions to recurrent maternal speech. *Infant Behavior & Development* 17: 159–164.
Dore, J. 1974. A pragmatic description of early language development. *Journal of Psycholinguistic Research* 3(4): 342–350.
Dore, J. 1975. Holophrases, speech acts, and language universals. *Journal of Child Language* 2: 21–40.
Edlund, J., D. House, and G. Skantze 2005. The effects of prosodic features on the interpretation of clarification ellipses. In *Proceedings of Interspeech-2005*, Lisbon, Portugal, 2389–2392.
Eefting, W. 1991. The effect of 'information value' and 'accentuation' on the duration of Dutch words, syllables and segments. *The Journal of the Acoustical Society of America* 89(1): 412–423.
Ernst, G. 1994. 'Talking circle': Conversation and negotiation in the ESL classroom. *TESOL Quarterly* 28(2): 293–322.
Ferguson, C. 1964. Baby talk in six languages. *American Anthropologist* 66: 103–114.
Ferguson, C. 1977. Baby talk as a simplified register. In *Talking to children, language input and acquisition*, ed. C. Snow and C. Ferguson. Cambridge: Cambridge University Press.
Fernald, A. 1989. Intonation and communicative intent in mother's speech to infants. Is the melody the message? *Child Development* 60: 1497–1510.
Fernald, A. 1992. Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. In *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J.H. Barkow, L. Cosmides, and J. Tooby, 391–428. Oxford: Oxford University Press.
Fernald, A., and P. Kuhl. 1987. Acoustic determinants of infant preference for motherese speech. *Infant Behavior & Development* 10: 279–293.
Fernald, A., and H. Morikawa. 1993. Common themes and cultural variations in Japanese and American mothers' speech to infants. *Phonetica* 57: 242–254.
Fox, A. 2001. *Prosodic features and prosodic structure. The phonology of suprasegmentals*. Oxford: Oxford University Press.
Frota, S. 2002. Nuclear falls and rises in European Portuguese: A phonological analysis of declarative and question intonation. *Probus* 14: 113–146.
Garnica, O.K. 1977. Some prosodic and paralinguistic features of speech to young children. In *Talking to children, language input and acquisition*, ed. C. Snow and C. Ferguson, 63–89. Cambridge: Cambridge University Press.
Gazdar, G. 1981. Speech act assignment. In *Elements of discourse understanding*, ed. A.K. Joshi, B.L. Webber, and I.A. Sag, 64–83. Cambridge: Cambridge University Press.
Geis, M.L. 1998. *Speech acts and conversational interaction*. Cambridge: Cambridge University Press.
Geluykens, R. 1987. Intonation and speech act type an experimental approach to rising intonation in queclaratives. *In Journal of Pragmatics* 11: 483–494.
Goodenough-Trepagnier, T., and F. Smith. 1977. Thematization and intonation in the organization of sentences. *Language and Speech* 20: 99–107.
Graesser, A.C., K. VanLehn, C. Rose, P. Jordan, and D. Harter. 2001. Intelligent tutoring systems with conversational dialogue. *AI Magazine* 22: 39–51.
Halliday, M.A.K. 1967. *Intonation and grammar in British English*. The Hague: Mouton.
Halliday, M.A.K. 1970. *A course in spoken English: Intonation*. Oxford: Oxford University Press.
Halliday, M.A.K. 1975. *Learning How to mean. Explorations in the development of language*. London: Edward Arnold.
Halliday, M.A.K. 1985/1994. *An introduction to functional grammar*. London: Edward Arnold.
Hasan, R. 1996. Semantic networks: A tool for the analysis of meaning. In *Ways of saying: Ways of meaning. Selected papers of Ruqaiya Hasan*, ed. C. Cloran, D. Butt, and G. Williams, 105–131. London: Cassell.

Johnstone, T., and K.R. Scherer. 2000. Vocal communication of emotion. In *Handbook of emotions*, 2nd ed, ed. M. Lewis and J. Haviland, 220–235. New York: Guilford Press.
Kloth, S., P. Janssen, F. Kraaitmat, and G.J. Brutten. 1998. Communicative styles of mothers interacting with their preschool-age children: A factor analytic study. *Journal of Child Language* 25: 149–168.
Ladd, D.R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.
Leech, G.N. 1983. *Principles of pragmatics*. London: Longman.
Levinson, S.C. 1994. *Pragmatics*. Cambridge: Cambridge University Press.
Levis, J.M. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33(1): 37–65.
Llinares-García, A. 2002. *La interacción lingüística en el aula de segundas lenguas en edades tempranas: análisis de un corpus desde una perspectiva funcional*. Tesis Doctoral, Universidad Autónoma de Madrid.
Long, M.H., and C.J. Sato. 1983. Classroom foreigner talk discourse: Forms and functions of teachers' questions. In *Classroom oriented research in second language acquisition*, ed. H. Seliger and M.H. Long, 268–287. Cambridge: Newbury House Publishers.
Martin, J.R. 1992. *English text: System and structure*. Amsterdam: John Benjamins.
McCarthy, J. 1998. Elephant 2000: A programming language based on speech acts. Stanford University. http://www-formal.stanford.edu/jmc/elephant/elephant.html
Mochizuki-Sudo, M. 1991. Production and perception of stress-related durational patterns Japanese learners of English. *Journal of Phonetics* 19: 231–248.
Morgan, B. 1997. Identity and intonation: Linking dynamic processes in an ESL classroom. *TESOL Quarterly* 31(3): 431–451.
Nagata, M., and T. Morimoto 1994. First steps towards statistical modelling of dialogue to predict the speech act type of the next utterance. In *Selected papers presented at the international symposium on spoken dialogue,* 193–203. New York: Elsevier North-Holland, Inc. Available from http://portal.acm.org/citation.cfm?id=208248.205964
Nooteboom, S.G., and J.G. Kruyt. 1987. Accents, focus distribution and the perceived distribution of given and new information: An experiment. *Journal of the Acoustic Society of America* 82(5): 1512–1524.
O'Connor, J.D., and G.F. Arnold. 1961. *Intonation of colloquial English*. London: Longman.
Palmer, H.E. 1922. *English intonation (with systematic exercises)*. Cambridge: Heffer.
Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35(2): 233–257.
Pike, K.L. 1945. *The intonation of American English*. Ann Arbor: University Michigan Press.
Radden, G. 1992. The cognitive approach to natural language. In *Thirty years of linguistic evolution*, ed. M. Pütz, 513–541. Philadelphia: John Benjamins.
Ramírez Verdugo, M.D. 2003. Análisis contrastivo de los sistemas entonativos del inglés en la interlengua de hablantes no nativos. Implicación en la organización de la información desde la perspectiva funcional. Estudio basado en un corpus computerizado de aprendices españoles de lengua inglesa. Tesis Doctoral, Universidad Autónoma de Madrid.
Riesco-Bernier, S. 2003. *From Discourse-Semantics to Phonology: A Cross-stratal Analysis of the Interpersonal Function in the EFL Pre-school Teacher Talk*. Trabajo de Suficiencia Investigadora, Universidad Autónoma de Madrid.
Riesco-Bernier, S., and J. Romero-Trillo. 2008a. The acoustics of 'Newness' and its pragmatic implications in classroom discourse. *Journal of Pragmatics* 40: 1103–1116.
Riesco-Bernier, S., and J. Romero-Trillo. 2008b. What does the tonic say in pre-school teacher talk in the EFL classroom? An acoustic-based analysis of tonicity. In *Institutional discourse in cross-cultural contexts*, Lincom studies in pragmatics, vol. 14, ed. R. Geluykens and B. Kraft, 147–169. München: Lincom Europa.
Rodríguez K.J., and D. Schlangen 2004. Forma, entonación y función de solicitudes de clarificación en diálogos instruccionales en alemán. *Revista de Procesamiento del Lenguaje Natural* 33: July.
Romero-Trillo, J., and A. Llinares-García. 2001. Communicative constraints in native/non-native preschool settings. *International Journal of Corpus Linguistics* 6(1): 27–46.

Rosch, E. 1977. Human categorization. In *Studies in cross-cultural psychology*, vol. 1, ed. N. Warren, 3–49. London: Academic.
Rosch, E. 1978. Principles of categorization. In *Cognition and categorization*, ed. E. Rosch and B. Lloyd, 27–48. Hillsdale: Lawrence Erlbaum.
Sachs, J. 1977. The adaptive significance of linguistic input to prelinguistic infants. In *Talking to children. Language input and acquisition*, ed. C. Snow and A. Ferguson, 51–61. Cambridge: Cambridge University Press.
Salaberri Ramiro, M.S. 1999. *El discurso del profesor en el aula y su relación con las tareas de aprendizaje.* Universidad de Almería: Servicio de Publicaciones. CD-Rom Tesis Doctoral 37.
Scott, M. 1998. *Wordsmith tools*. Version 3.0. Oxford: Oxford University Press.
Searle, J. 1969. *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
Searle, J. 1975. Indirect speech acts. In *Syntax and semantics 3: Speech acts*, ed. P. Cole and J.L. Morgan, 59–82. New York: Academic.
Searle, J. 1976. The classification of illocutionary acts. *Language in Society* 5: 1–24.
Sinclair, J.Mc.H., and D. Brazil. 1982. *Teacher talk*. Oxford: Oxford University Press.
Sinclair, J.Mc.H., and M. Coulthard. 1975. *Towards an analysis of discourse. The English used by teachers and pupils*. London: Oxford University Press.
Sluijter, A.M.C., and J.M.B. Terken. 1993. Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica* 50: 180–188.
Snow, C. 1979. Conversations with children. In *Language acquisition*, ed. P. Fletcher and M. Garman, 363–376. Cambridge: Cambridge University Press.
Stern, D.N., S. Spieker, R.K. Barnett, and K. MacKain. 1983. The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language* 10: 1–15.
Sullivan, J., and F.D. Horowitz. 1983. The effects of intonation on infant attention: The role of the rising intonation contour. *Journal of Child Language* 10: 521–534.
Swerts, M., and R. Geluykens. 1993. The prosody of information units in spontaneous monologue. *Phonetica* 50: 189–196.
Tench, P. 1996. *The intonation systems of English*. London: Cassell.
Thompson, W.F., E.G. Schellenberg, and G. Husain. 2004. Decoding speech prosody: Do music lessons help? *Emotion* 4: 46–64.
Trainor, L.J., C.M. Austin, and R.N. Desjardins. 2000. Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science* 11: 188–195.
Tyler, A. 1992. Discourse structure and the perception of incoherence in international teaching assistants' spoken discourse. *TESOL Quarterly* 26(4): 713–729.
Verschueren, J. 1999. *Understanding pragmatics*. New York: Arnold.

Online Resources

http://esl.about.com/od/speakingadvanced/a/timestress.htm
www.englishclub.com/pronunciation/
www.phon.ucl.ac.uk/resource/tutorials.html
www.qwertystudios.com/speech/tts-study/study-accurate-pronunciation/intonation-skills.html
www.soundsofenglish.org/pronunciation/index.html

Chapter 12
The Pragmatic Function of Intonation: Cueing Agreement and Disagreement in Spoken English Discourse and Implications for ELT

Lucy Pickering, Guiling (Gloria) Hu, and Amanda Baker

12.1 Introduction

Approaches to spoken discourse analysis have demonstrated that prosodic features in English such as intonation, stress and pausing play a key role in determining how participants manage interaction (Akker and Cutler 2003; Brazil 1997; Chafe 1994; Couper-Kuhlen 1996; Cutler et al. 1997). These features are particularly significant when considering the discourse-pragmatic functions of intonation (Chun 1988), where prosody has been shown to form a natural link between linguistic and sociolinguistic aspects of language (Brazil 1997; Gumperz 1982).

Non-referential functions of pitch variation include regulation of turn-taking in conversation and the communication of sociolinguistic information such as status differences, solidarity or social distance between interlocutors (Couper-Kuhlen and Selting 1996).

Despite its important role, prosody has traditionally been neglected in cross-cultural studies of pragmatics and is rarely approached in English language teaching (ELT) literature (although see Cauldwell 2001; Levis 1999); yet the small body of existing research suggests that there may be a mismatch of prosodic cues in second language (L2) learners' expression of (dis)agreement, which may be detrimental to their interactions with native English speaker interlocutors (Hewings

L. Pickering (✉)
Department of Literature and Languages, Texas A&M-Commerce, Commerce, TX, USA
e-mail: lucy_pickering@tamu-commerce.edu

Guiling (Gloria) Hu
National Foreign Language Center, University of Maryland, College Park, MD, USA

A. Baker
Faculty of Education, University of Wollongong,
Wollongong, Australia

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_12,

1995; Pickering 1999, 2004). In light of these findings, we investigate the pragmatic function of intonation in cueing (dis)agreement in the naturally-occurring discourse of American English speakers and Chinese learners of English. We are particularly interested in the possible role of pitch level matching between interlocutors to cue (dis)agreement.

12.2 Literature Review

Much of the foundational work on the description of sequences of agreement and disagreement in English comes from the area of conversational analysis, where the focus has been on the sequential organization of conversation and the examination of turn-taking structures (Couper-Kuhlen and Selting 1996). This research has established that there is a strong preference for agreement between interlocutors (Davidson 1984; Pomerantz 1984; Sacks 1987); thus, sequences are generally "structured so as to maximize occurrences of stated agreements and disagreement turns/sequences so as to minimize occurrences of stated disagreements." (Pomerantz 1984, p. 64). Preferred options include a 'minimization of the gap' between speaker turns in which the second speaker is invited to agree with the assessment made by the first speaker:

(1)
A: well that was fun Claire
B: Yeah, I enjoyed every minute of it (p.60)

As the dispreferred option, disagreements may be prefaced with initial agreement components in order to 'downgrade' disagreement:

(2)
A: You are afraid of your father
B: Oh yes. Definitely. I- I am. To a certain extent.
(Sacks 1987, p. 63)

Or speakers may formulate their question in such a way that disagreement will be avoided:

(3)
A: Those're- Are those that same- No that's not the present I gave you
B: No I know- I've broken from the pattern
(Sacks 1987, p. 64)

These examples also show additional strategies used to mitigate disagreement, including prefacers such as 'uh' and 'well', delay devices including 'repair initiators' such as 'what' or 'hmmm?' or silence, i.e., an overlong pause before turn initiation.

In addition to lexical and syntactic devices, (dis)agreement options may also be cued by phonetic features, including speakers' choice of pitch (Ogden 2006).

A consistent feature of agreement sequences noted in the literature is variously known as 'melodic matching' (Couper-Kuhlen 1996), 'pitch concord' (Brazil 1997) and 'prosodic matching' (Szczepek Reed 2006) and refers to a preference for a second speaker to match his/her initial pitch choice in terms of relative pitch height to the final pitch choice of the first speaker. In contrast, a mismatch of pitch choice or 'concord breaking' (also referred to as 'prosodic non-matching' by Szczepek Reed 2006) can mark dissonance when a significantly higher or lower pitch choice is used by the second speaker.

Both French and Local (1983) and Wennerstrom (2001) found that in interruptions and other instances of competition for the floor, speakers raised both their pitch and volume; Wennerstom proposes that "for English speakers, a high key response can convey a contrast in attitude with respect to the prior contribution" (p. 240). Selting (1996) and Gunthner (1996) note the same phenomenon in German conversational data, where mismatched high key responses cued rebukes or amazement and required repair.

Significantly lower pitch choices by a second speaker resulting in a mismatch between interlocutors can also signal a discrepancy or discord between speakers in English (Schegloff 1998; Wennerstrom 2001). Muller (1996) reports similar findings in the use of recipiency tokens such as 'uh huh', 'yeah' and 'right' in Italian. While affiliating tokens were prosodically matched with the emerging talk, disaffiliating tokens exhibited concord breaking, as they were realized with a significantly lower pitch register.

In addition to pitch concord, i.e., interlocutors' matching of pitch levels in consecutive utterances, analysts have also looked at pitch movement or the shape of pitch contours in the assessment of speaker contributions. Using Brazil's (1985/1997) model of intonation in discourse as a framework, Hewings (1995) reports that English-speaking informants uniformly used a rising tone when contradicting a previous speaker in order to avoid the appearance of overt disagreement that might be inferred from a falling tone. Rising tones also co-occurred with speakers' choices to withhold agreement. Hewings concludes that there is an "exploitation of the Rising/Falling opposition for socially integrative purposes" (p. 262). In an analysis of teacher-student exchanges, Pickering (2001) also found that teachers exploited tone choices in order to promote social convergence in the classroom, particularly when it came to disagreeing with a student response. Teachers consistently used a rising tone to indicate withholding of agreement, which communicated to the student that the answer was incorrect. A similar 'yes, but' strategy found to be communicated phonetically in classroom discourse is a withholding of agreement in the form of a level tone on delay devices such as //➔ WELL// or //➔UM//. This use of prosodically significant lexical continuers is described by Muller (1996 p.133) as "short tokens, 'long' prosody."

Ogden (2006) looks at both tonal contours and pitch concord in the production of second assessments in (dis)agreement. In cases of strong agreement he describes a phonetic "upgrade" that comprises an expanded pitch range, a higher pitch in the speaker's range and the use of more dynamic pitch contours. Similar features co-occurred with overt disagreements, although use of this option was rare.

More typically, disagreements were prefaced with an agreement marker such as a lexical continuer and demonstrated a 'phonetic downgrade', comprising a narrower pitch range, a lower pitch and a lack of dynamic pitch movement.

In both English and German agreement sequences, Koester (1990) found that pitch matching in speakers' mid-range was most common. However, low pitch and high pitch concord-breaking responses did occur. Tonal contours were also varied in second assessments and no particular tone (rising, falling or level) was found to be more prevalent. Koester's data show very few disagreements between speakers and no consistent intonational features were found for either German or English, although English speakers preferred to use a rising tone for initial agreement markers in agree + disagree (i.e., 'yes, but') sequences while German speakers preferred a level tone. Overall, findings suggest that agreement sequences between interlocutors may be supported by some kind of "prosodic alignment" (Szczepek Reed 2006, p.60) between speakers while disagreement sequences may exhibit prosodic disaffiliation.

12.2.1 The Prosody of L2 (Dis)agreement Sequences

There is strikingly little research on the prosodic characteristics of learner language in general and this includes investigation of the prosodic features of (dis)agreement sequences. With regard to pitch concord, Koester (1990) found that a lack of pitch concord between German learners of English (in this case use of a low pitch choice where a mid pitch choice was expected) prompted a first speaker to confirm their partner's agreement. This suggests that the L2 speakers understood the function of concord breaking in this case and perceived it as a meaningful pragmatic cue. Similar results were found in Pickering (2009) in an investigation of intonation as a pragmatic resource in English as a Lingua Franca (ELF) interactions. Although not focused specifically on (dis)agreement sequences, data showed that pitch level choices and the shape of tonal contours were used by interlocutors to signal trouble spots and negotiate their resolution.

On the other hand, data comparisons between English native speakers (NSs) and L2 speakers suggest that there may be significant differences between the two groups in their use of pitch cues to signal pragmatic intent. Hewings (1995) found that advanced learners of English from Korea, Greece and Indonesia showed a tendency toward using falling tones in disagreement sequences whereas NSs consistently used rising tones when contradicting a previous speaker, to avoid the appearance of overt disagreement implicit in a falling tone. Similar results were found by Ramírez Verdugo (2005) with Spanish learners of English who used primarily falling tones and thus, did not "express the reservation implied in the native speakers' fall-rise contour" (p. 2100).

Mennen (2007) finds that there are some significant differences in pitch range characteristics between native English speakers and German speakers of English. She suggests that German exhibits a narrower pitch range than English, thus German

speakers may transfer this characteristic to English and be perceived as more negative. Such differences may also result in unintentional concord breaking by L2 speakers of English. Anderson (1990) reports an interaction between a NS of English and a Dutch speaker of English in which high pitch choices by the NNS project conflict and result in a failed interaction. Pickering (2002, pp. 11–12) reports a similar confusion over interpretation of a Chinese speaker's pitch choices, which confounds the expectations of a North American undergraduate student and results in miscommunication.

The data we investigate here focus on Chinese learners of English (CLsE) and in light of the possible impact of cross-linguistic transfer, we were also interested in the prosodic characteristics of (dis)agreement sequences in Chinese. To date, there are few studies focusing on characterizing pragmatic competence of NSs of Chinese and none that consider the possible role of prosodic cues in the manifestation of (dis)agreement in spoken discourse. Recently, however, researchers have begun to examine possible attitudinal functions of Chinese intonation. In a series of studies investigating friendly speech in Mandarin, Li and associates (Chen et al. 2004; Li et al. 2004; Li and Wang 2004) found that the average pitch mean was higher in friendly speech than in neutral speech. Hu (2005) found that register-raising is also used to show surprise. Yuan et al. (2002) further report that the pitch used to express anger, fear or joy is higher than that used to express sadness. In addition, they suggest the entire pitch contour fluctuates more greatly when expressing anger and joy as opposed to fear and sadness. Consideration of these studies as a whole suggests that Chinese speakers may use a higher pitch register and a greater contour fluctuation to express an attitude that is not neutral.

In this study, we extend the current research by investigating (dis)agreement sequences in native speakers of American English and Chinese learners of English. We focus specifically on use of pitch concord, namely, "a preferential relationship holding between pitch level choices in adjacent utterances" (Anderson 1990, p.106) as a cue to signal (dis)agreement between interlocutors.

12.3 Method

12.3.1 Participants

Twelve native speakers of American English (NSE) and 12 Chinese learners of English (CLsE) participated in the project. Both groups comprised undergraduate and graduate students enrolled in a tertiary institution in the South Eastern United States. Six male (M) and six female (F) participants in each group formed two male-male, two male–female and two female-female pairs. As Liang and Jing (2005) found that rates of disagreement between Chinese speakers decreased with an increase in social distance, particular care was taken to choose pairs of speakers who were familiar with each other (e.g., colleagues and friends in the same program) and who had equal social status. This resulted in equitable participation by

Table 12.1 Chinese learners of English

	Mean	Range
Age	27	23–31
TOEFL	616	590–650
Age at beginning of English instruction	12	10–13
Years of formal instruction	12	10–15
Years of residence in the US	3;6	1–5
Self-evaluation of English proficiency		
Speaking	5.8	5–7
Listening	7.4	5–8
Reading	7.6	6–9
Writing	6.3	5–8

individual speakers in the interactions (Kasper 2000). The Chinese learners of English were administered a questionnaire prior to the data collection, which included a self-evaluated proficiency score on a 10-point scale (10 represented NS competence and 1 represented no experience with English.) Their responses are given above in Table 12.1.

In total, we collected approximately 23 min of data from the six native speaker pairs and approximately 35 min of data from the Chinese learners of English.

12.3.2 Procedures

Pairs of speakers were seated next to each other in a quiet room in front of a laptop computer. Each speaker wore a Telex SCHF745 headset microphone and was recorded using a Telex FMR-150C wireless system and a Sony TCD-D8 Digital Audio Tape-recorder (DAT). In an adaptation of the method used by Koester (1990) to elicit (dis)agreement sequences, speakers were shown a series of pictures of ten concept cars and asked to come to a mutual agreement as to their favorite car (see Appendix). The participants controlled the laptop and viewed the cars in any order they preferred. Each conversation was transcribed verbatim and in its entirety. These transcripts were read by six native speakers of English who marked places in the transcripts where they identified (dis)agreement sequences. Instances of (dis)agreement that were marked by four out of six of the judges (i.e., more than 70% of the judges) were analyzed for pitch structure. Written transcripts were used for this identification in order to avoid a circular identification of (dis)agreement pitch patterns. Previous research suggests that speakers use multiple cues across linguistic systems to indicate pragmatic intent (Pickering 2001, 2004; Tyler 1992; Tyler and Bro 1993); thus, we anticipated that sequences primarily identified by syntactic or lexical cues by our judges would also exhibit some consistency in intonational cues.

12.3.3 Data Analysis

DAT recordings of (dis)agreement sequences were transferred to a Kay Pentax 4500 Computerized Speech Laboratory (CSL). Fundamental frequency (F_0) traces and spectrograms were generated for all the data using the relevant functions of the CSL. All data were subject to both auditory and instrumental analysis (Pickering 2001). Analysis focused on the identification of pitch level choices in adjacent utterances by each speaker in a pair, i.e., evidence of the operation of pitch (dis)concord between interlocutors to cue (dis)agreement. Our definition of pitch concord derives from Brazil's discourse intonation model, in which the final prominent pitch choice of one turn is compared to the first prominent syllable of the consecutive turn.

A comparison of pitch concord patterns across multiple voices, particularly if participants are both male and female, requires raw frequency values (Hz) to be converted to a relative scale. To achieve this, we followed the procedure used by Couper-Kuhlen (1996) to analyze data on pitch matching by converting each measurement to semitone (ST) values using a formula developed by t'Hart et al. (1990, p. 24): "Hz values are recalculated on a semitone scale relative to each voice range and expressed as ST intervals from the lowest Hz value a given speaker is inclined to use" (p. 374).

Following Couper-Kuhlen, the baseline for each speaker was established through measurement of all their recorded utterances. Raw Hz values were converted to ST values for each speaker and the difference in STs in consecutive utterances between speakers was recorded. Although Couper-Kuhlen (1996) does not specify an exact cut-off point for what comprises a pitch match,[1] her data examples suggest that pitch values less than or equal to 1 ST constitute pitch matching between consecutive utterances by different speakers (see, for example, p. 376). She also notes, however, that there are different degrees of matching and 'modified matches' may be less precise (see, for example, p. 378). For this reason, we have also included matches that are less than or equal to 2 STs as a separate category.

12.4 Results

12.4.1 Native Speakers of American English

The transcripts of the six NS-NS pairs yielded 76 (dis)agreement sequences with a heavy bias against the dispreferred option of disagreement: 68 agreement sequences and 8 disagreement sequences. As noted earlier, (dis)agreement sequences were identified from the written transcripts of the interactions between

[1] It should also be noted that Couper-Kuhlen investigated quoting and mimicry rather than agreement sequences.

Table 12.2 Pitch concord analysis for NS-NS agreement sequences

Transcripts	# of agreements overall	# of consecutive pitch choices less than or equal to one semitone apart (≤1 ST)	# of consecutive pitch choices less than or equal to two semitones apart (≤2 STs)	# of consecutive pitch choices more than two semitones apart (>2 STs)
M8-M9	20	11	4	5
F10-M10	8	2	4	2
F6-M4	17	6	3	8
M11-M12	7	4	2	1
F11-F12	6	4	0	2
F2-F3	10	5	3	1
Totals	68	32	16	19

participants; thus, there was a preference to identify sequences that could be clearly recognized based on lexical and syntactic cues. This resulted in a preference for the identification of short assessment pairs with overt lexical cues such as those shown in examples 4 & 5:

(4) Agreement
M9: Somebody has very expensive taste
M8: Yeah, no kidding
(5) Disagreement
F11: I kind of like that one
F12: Umm no, I don't really like that one

No significant differences were found between the prosodic characteristics used by male and female speakers and no further distinctions were drawn between the two groups.[2]

12.4.1.1 Agreement Sequences

The results of the pitch concord analysis for agreement sequences are shown in Table 12.2.

With regard to agreement sequences, 48% of the sequences demonstrated matching in the form of pitch concord between consecutive utterances by two speakers at ≤1 ST as shown in Fig. 12.1.

When pitch matching between speakers was defined less strictly as ≤2 STs, instances of pitch concord increased to 72%. An example is shown in Fig. 12.2.

[2] See also Rees-Miller (2000) for similar findings regarding rate of disagreement and use of mitigating devices in relation to gender.

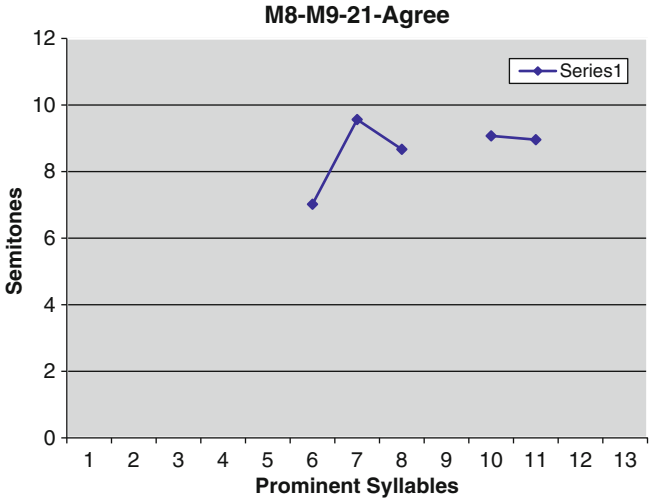


Fig. 12.1 M8: //YEAH //i MEAN it's the OPposite//
M9: //YEAH// it's TRUE//

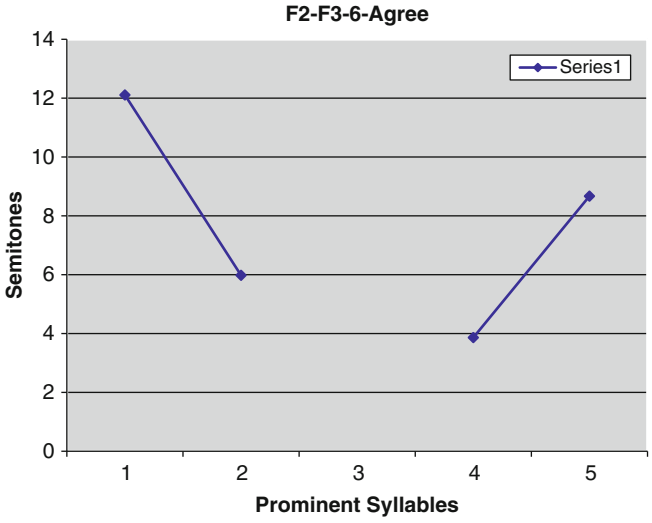


Fig. 12.2 F2: //that's like a car tom CRUISE would DRIVE//
F3: //in that MOVie// YEAH//

The remaining cases fell within the third group of >2 STs, which was not considered to mark pitch concord between speakers. Examination of these 19 sequences revealed additional types of pitch matching behaviors that have also been identified in the literature as cueing agreement between speakers and which may substitute in these cases for pitch concord. The most common were instances of a high key

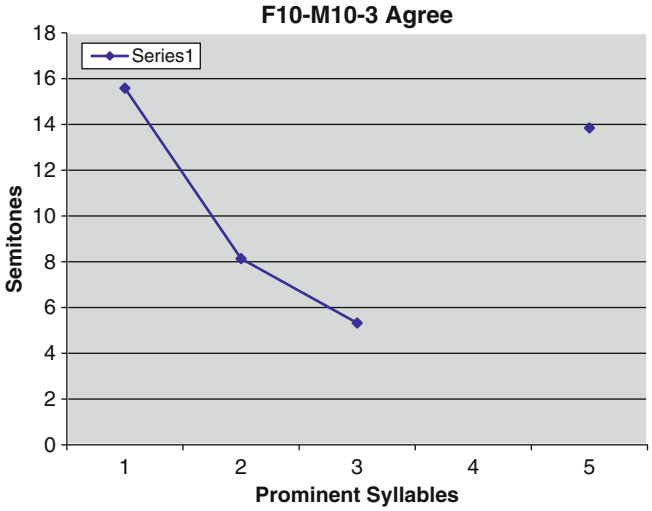


Fig. 12.3 F10: //THAT looks like a PEAnut on WHEELS!//
M10: //YEAH!//

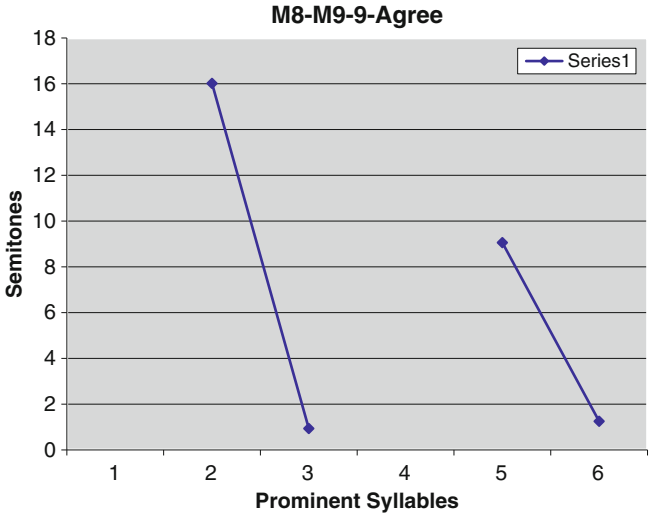


Fig. 12.4 M8: //ALL of these are pretty high-end CARS//
M9: //they're ALL very high end AREN'T they//

response by the second speaker which Koester (1990, p. 86) describes as "particularly enthusiastic agreement" and is shown in Fig. 12.3.

Two additional sequences manifested *actual* as opposed to *relative* pitch matching and two agreement sequences exhibited pitch contour matching, as shown in Fig. 12.4.

Table 12.3 Pitch concord analysis for NS-NS disagreement sequences

Transcripts	# of disagreements overall	Distance in STs between consecutive pitch choices	Pitch of second utterance
M8-M9	1	17	Higher (M)
F10-M10	1	7	Higher (F)
F6-M4	1	13	Lower (M)
F11-F12	2	7	Lower (F)
		14.6	Higher (F)
F2-F3	3	11	Higher (F)
		15	Higher (F)
		10.6	Lower (F)
Total	8		

Following this further analysis, only nine agreement sequences (13%) could not be shown to demonstrate any transparent relationship between consecutive utterances by separate speakers and mutual pitch choices.

12.4.1.2 Disagreement Sequences

The results of the pitch analysis for disagreement sequences are shown in Table 12.3.

The disagreement sequence data revealed that interlocutors consistently signaled their lack of agreement with the previous utterance with a discordant pitch choice in addition to lexical and syntactic cues. An example is shown in Fig. 12.5.

As Table 12.3 shows, pitch choices were either significantly lower or higher in the second utterance and choice was not dictated by the gender of the speaker. In all cases, consecutive pitch choices between speakers were separated by large distances in terms of STs (mean = 11.9 STs), which were far greater than those found between agreement sequences.

In sum, pitch concordance analysis across consecutive utterances between two NSs revealed that while pitch concord may be used as a cue to signal agreement between interlocutors, it is not a consistent feature of agreement sequences. In disagreement sequences, however, the second NS interlocutor consistently signaled disagreement with a discordant pitch choice, suggesting that this may be a considerably stronger discourse cue.

12.4.2 Chinese Learners of English

Ratings of the six NNS-NNS transcripts of Chinese learners of English (CLsE) produced 69 (dis)agreement sequences: 56 agreement sequences and 13 disagreement sequences. These results were highly consistent with the NS data, both in

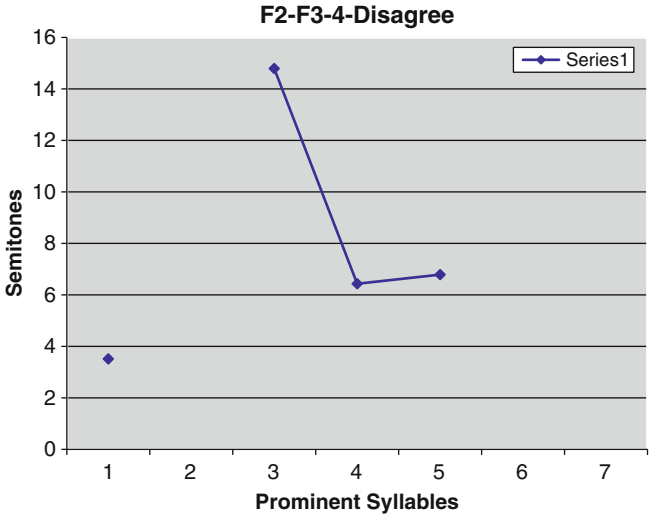


Fig. 12.5 F2: //That's oK//
F3: //WELL// YEAH// that's BETter//

terms of numbers of instances and in the nature of (dis)agreement sequences. As with the NS-NS transcripts, raters tended to agree most often on short assessment pairs, as shown in Examples 6 and 7:

(6) Agreement
F4: Not good
F5: Yeah, not good
(7) Disagreement
M7: I like the color
F7: I don't really like the color

12.4.2.1 Agreement Sequences

The results of the pitch concord analysis for agreement sequences are shown in Table 12.4.

Forty-one percent of agreement sequences in the CLsE data exhibited pitch matching at ≤1 ST; when extended to ≤2 STs; this accounted for 77% of sequences and is directly comparable to the findings for the NS data. The remaining 13 cases were again examined for evidence of additional pitch devices. Six instances of similar kinds of pitch matching to that found in the NS data were found, comprising two examples of enthusiastic agreement, two examples of matching pitch contours and lexis (see Fig. 12.6); and two examples of absolute pitch matching. Following this

Table 12.4 Pitch concord analysis for the CLsE agreement sequences

Transcripts	# of agreements overall	# of consecutive pitch choices less than or equal to one semitone apart (≤1 ST)	# of consecutive pitch choices less than or equal to two semitones apart (≤2 STs)	# of consecutive pitch choices more than two semitones apart (>2 STs)
M5-M6	7	1	4	2
F7-M7	10	4	1	4
F8-F9	9	6	4	0
F4-F5	10	5	5	0
M2-M3	8	3	4	1
M1-F1	12	4	2	6
Totals	56	23	20	13

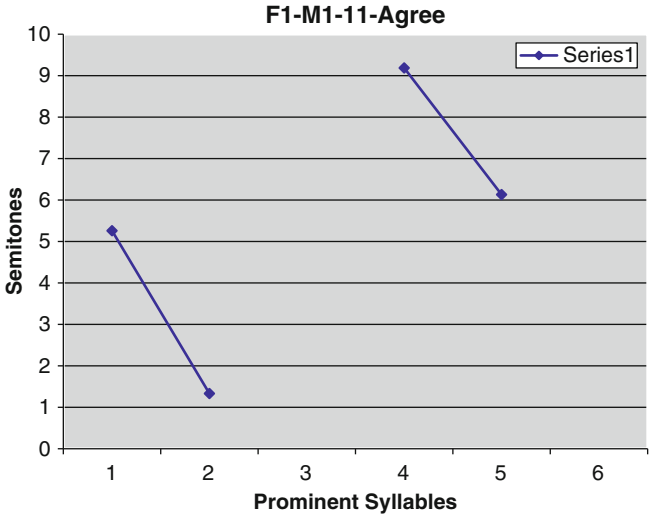


Fig. 12.6 F1: //this CAR is ELegant//
M1: //ummm this CAR is ELegant//

additional analysis, seven cases (12%) of agreements could not be accounted for by any pitch related phenomena.

12.4.2.2 Disagreement Sequences

The results of the pitch analysis for disagreement sequences are shown in Table 12.5.

These data exhibited considerably less consistency with regard to pitch discord in comparison to the NS data. Most notably, disagreements were not uniformly

Table 12.5 Pitch concord analysis for CLsE disagreement sequences

Transcripts	# of Disagreements overall	Distance in STs between consecutive pitch choices	Pitch of second utterance
M5-M6	1	10	Higher (M)
F7-M7	2	4.2	Lower (F)
		9.1	Higher (F)
F8-F9	1	2	Higher (F)
		14.6	Higher (F)
F2-F3	3	11	Higher (F)
		15	Higher (F)
		10.6	Lower (F)
F4-F5	2	3.5	Lower (F)
		5	Higher (F)
F2-F3	1	2	Lower (F)
M1-F1	3	12.5	Lower (F)
		13	Higher (M)
		11.8	Higher (M)
Totals	13		

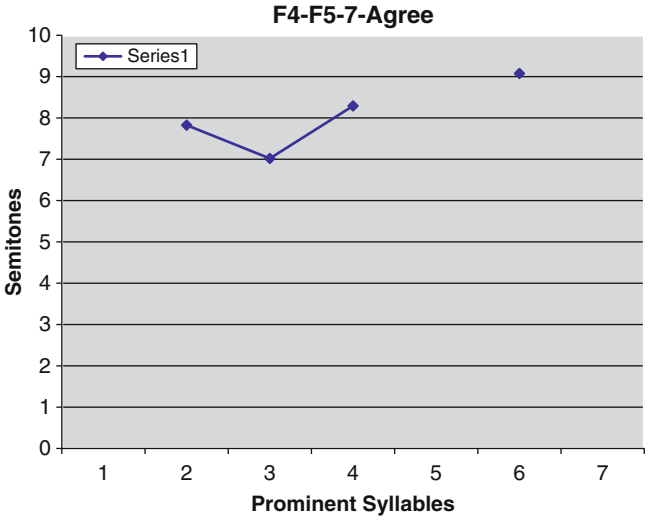


Fig. 12.7 F4: //SOMEthing WEIRD RIGHT//
F5: //YEAH//

signaled by CLsE with discordant pitch choices in second utterances. Unlike NS data, second utterances in disagreement sequences varied in distance from 2 STs to 15 STs. This is illustrated in Figs. 12.7 and 12.8, in which both second utterances, one in agreement and one in disagreement with the previous utterances, show similar variability in terms of distance in pitch from the first utterance.

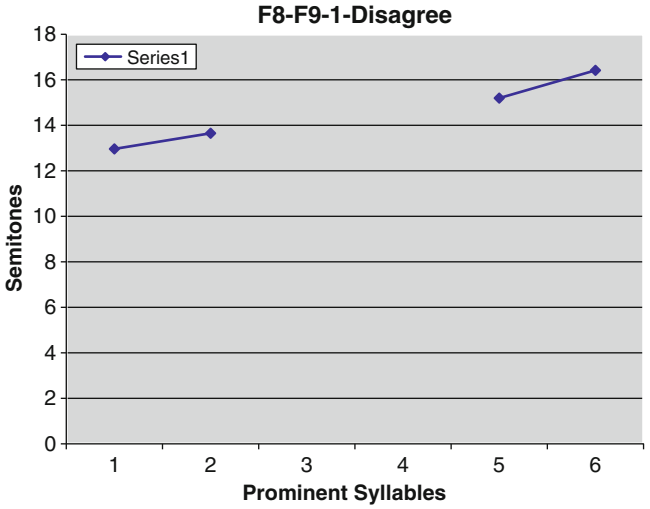


Fig. 12.8 F8: //THIS pretty COOL//
F9: //BUT I prefer the SILver//

12.5 Discussion

In an investigation of the role of prosodic cues in (dis)agreement sequences in L1 and L2 spoken discourse, we examined evidence of pitch concord in both NS-NS and NNS-NNS interactions. In the majority of cases, both NSs and NNSs manifested pitch concord in agreement sequences; that is, a relationship of ≤2 semitones pertained between the speakers' utterances. Instances that did not show pitch concord often demonstrated additional types of prosodic matching, such as matching pitch contours.

Neither group, however, showed uniform use of pitch matching in agreement sequences; thus it was not a necessary condition. This is anticipated under a discourse intonation model. As Brazil points out, there is no absolute requirement that a speaker obey constraints such as the concord principle. Rather, the intonation system operates on the Gricean cooperative principle (Grice 1989) that generally speaking, speakers' contributions are designed to be understood. If pitch matching is a conventional cue for agreement, a preference for pitch concord would be expected and this is what these data show.

The most consistent finding in the NS data was the use of discordant pitch choices (either significantly lower or higher) in disagreement sequences. This suggests that discordant pitch may be a very robust discourse cue in native speaker interaction. Szczepek Reed (2006, p. 77) proposes that prosodic non-matching in this case is "an iconic representation of the non-matching of opinions between two participants." This finding did not hold in the NNS data and this could be a possible area for pedagogical intervention. Previous research has demonstrated that NNSs may display an

overall narrower pitch range regardless of L1 (Mennen 1998; Pickering 2004), thus this is not necessarily a feature of L1 transfer (note that the literature cited in 2.1.1. suggests that CLsE use pitch register raising and pitch contour fluctuations in their L1) but rather L2 development.[3]

The analyses also raised a number of important methodological questions that need to be addressed; most crucially, how pitch concord should be operationalized. No absolute value for what qualifies as pitch matching is given in the earlier studies that we draw from. In addition, while we chose to consider only prominent syllables as salient cues for pitch measurement, previous work has also included non-prominent pitch matching (e.g., Couper-Kuhlen 1996). While it is clearly important to establish relative pitch ranges in order to plot pitch matching, baselines are estimated and there is the possibility of measurement error.[4] Largely for this reason, we noted pitch concord patterns up to and including two semitones as possible pitch matches.

12.6 Pedagogical Implications

Davies (2004) states that the ability to successfully perform pragmatic functions such as (dis)agreement sequences is crucial for the development of interactional competence and has broad practical implications for second language teaching. Incorrect use or interpretation of a speech act in an unfamiliar culture will not only cause communication breakdowns but may also intensify misunderstandings between two cultures (Zhang 2001). In particular, if L1 hearers perceive an L2 speaker to have a high linguistic proficiency, misuse of a speech act is frequently not interpreted as a lack of communicative competence but a sign of an unpleasant personality (e.g., Tannen 1986). This is particularly true of a contributing linguistic system as tacit as prosody, where our impressions of speakers are likely to suffer based on "misperceptions and misplaced stereotypes" deriving from inappropriate use of intonation (Mennen 2007).

Yet, as Wrembel (2007) notes, despite a consensus regarding the significance of prosodic features for successful communication, "prosody still appears to be the 'problem child' from the pedagogical perspective" (p. 189). Certainly, a cursory review of ESL/EFL textbooks shows limited if any discussion of the role of prosodic features in face-threatening speech acts such as disagreement sequences. A reluctance to address the role of intonation in particular may have been hampered by its traditional representation as a "half-tamed savage",[5] lying on the edge of language and more appropriate for paralinguistic investigation. Current research trends

[3] Although see Mennen (2007) for further discussion of L1 transfer.

[4] Note that a speaker's baseline cannot be calculated automatically as an average of the lowest produced frequencies, as this will include instances of creaky voice among other voice quality issues.

[5] From Bolinger (1978), cited in Vaissiere (1995).

may also prioritize intelligibility in English as a Lingua Franca (ELF) interaction, in which the value of pitch movement as a feature of effective interaction has also been challenged (Jenkins 2000, 2002). However, such sentiments do not aid the NNS who interacts with NSs on a regular basis, such as the population of CLsE investigated here. In this situation, prosody contributes significantly to interactional competence and serves to establish a crucial collegial bond between speakers. Intonational features are necessary for successful communication and systematic attention to prosody in the English language classroom is key.

Jilka (2007) suggests that we might teach learners "conscious control" of features such as pitch range and also suggests the use of speech technology to facilitate this. This has also been proposed in applications of speech visualization technology to the teaching of ESL/EFL by Chun (1998) and Levis and Pickering (2004).

Acton (2010) has developed a haptic method for teaching intonation in which learners use both movement and touch to coordinate the body with prosodic and segmental features, with the intention of producing fluent and intelligible speech.

Davies (2004, pp. 225–227) also proposes a pedagogical plan to develop awareness of cross-cultural pragmatics, which includes the role of prosody and is grounded in four central principles: (1) a teaching focus on discourse as opposed to isolated acts; (2) developing learners' ability to identify patterns through discourse analysis; (3) an understanding of the unique cultural context of social norms; (4) an understanding of the uniqueness of each interaction as it emerges moment to moment. If each of these principles is applied directly to teaching intonation in the classroom, learners will have access to the information they need to successfully navigate the kind of pragmatic function we have focused on here.

Appendix

Pictures of Concept Cars

References

Acton, W. 2010. Full-bodied, systemic, multiple modality pronunciation instruction. Workshop given at the 44th Annual TESOL Convention, Boston, MA.

Akker, E., and A. Cutler. 2003. Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition* 6(2): 81–96.

Anderson, L. 1990. Intonation, turn-taking and dysfluency: Non-natives conversing. In *Papers in discourse intonation*, Discourse analysis monograph, vol. 16, ed. M. Hewings, 102–113. Birmingham: University of Birmingham: English Language Research.

Bolinger, D. 1978. Intonation across languages. In *Universals of human language: Phonology*, ed. J. Greenberg. Stanford: Stanford University Press.

Brazil, D. 1997. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

Cauldwell, R. 2001. *Streaming speech*. Birmingham: Speech in Action.

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

Chen, F., A. Li, H. Wang, T. Wang, and Q. Fang. 2004. *Acoustic analysis of friendly speech*. Beijing: Phonetics Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

Chun, D. 1988. The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal* 72: 295–303.

Chun, D. 1998. Signal analysis software for teaching discourse intonation. *Language Learning and Technology* 2: 61–77.

Couper-Kuhlen, E. 1996. The prosody of repetition: On quoting and mimicry. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 366–405. Cambridge: Cambridge University Press.

Couper-Kuhlen, E., and M. Selting, eds. 1996. *Prosody in conversation*. Cambridge: Cambridge University Press.

Cutler, A., D. Dahan, and W. van Donselaar. 1997. Prosody in the comprehension of spoken language: A literature review. *Language and Speech* 40: 141–201.

Davidson, J. 1984. Subsequent versions of invitations, offers, requests, and proposals dealing with potential or actual rejection. In *Structures of social action*, ed. J.M. Atkinson and J. Heritage, 102–128. Cambridge: Cambridge University Press.

Davies, C. 2004. Developing awareness of crosscultural pragmatics: The case of American/German sociable interaction. *Multilingua* 23: 207–231.

French, P., and J. Local. 1983. Turn-competitive incomings. *Journal of Pragmatics* 7: 17–38.

Grice, P.H. 1989. *Studies in the way of words*. Cambridge: Harvard University Press.

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

Gunthner, S. 1996. The prosodic contextualization of moral work: An analysis of reproaches in 'why'-formats. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 271–302. Cambridge: Cambridge University Press.

Hart't, J., R. Collier, and A. Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Hewings, M. 1995. Tone choice in the English intonation of non-native speakers. *IRAL* 33(3): 251–265.

Hu, F. 2005. *A phonetic study of prosody of Wh-words in standard Chinese*. Beijing: Phonetics Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

Jenkins, J. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.

Jenkins, J. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics* 23(1): 83–103.

Jilka, M. 2007. Different manifestations and perceptions of foreign accent in intonation. In *Non-native prosody: Phonetic description & teaching practice*, ed. J. Trouvain and U. Gut, 77–96. Berlin: Mouton de Gruyter.

Kasper, G. 2000. Data collection in pragmatics. In *Culturally speaking*, ed. H. Spencer-Oatey, 316–341. London/New York: Continuum.

Koester, A. 1990. The intonation of agreeing and disagreeing in English and German. In *Papers in discourse intonation*, Discourse analysis monograph, vol. 16, ed. M. Hewings, 83–101. Birmingham: University of Birmingham: English Language Research.

Levis, J. 1999. Intonation in theory and practice, revisited. *TESOL Quarterly* 33: 37–64.

Levis, J., and L. Pickering. 2004. Teaching intonation in discourse using speech visualization technology. *System* 32: 505–524.

Li, A., and H. Wang. 2004. *Friendly speech analysis and perception in standard Chinese*. Beijing: Phonetics Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

Li, A., F. Chen, H. Wang, and T. Wang. 2004. *Perception on synthesized friendly standard Chinese speech*. Beijing: Phonetics Laboratory, Institute of Linguistics, Chinese Academy of Social Sciences.

Liang, G., and H. Jing. 2005. A contrastive study on disagreement strategies for politeness between American English & Mandarin Chinese. *Asian EFL Journal* 7(1): 1–12.

Mennen, I. 1998. Can second language learners ever acquire the intonation of a second language? In *Proceedings of the ESCA workshop on speech technology in language learning*, Marholmen, Sweden.

Mennen, I. 2007. Phonological and phonetic influences in non-native intonation. In *Non-native prosody: Phonetic description & teaching practice*, ed. J. Trouvain and U. Gut, 53–76. Berlin: Mouton de Gruyter.

Muller, F. 1996. Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 131–176. Cambridge: Cambridge University Press.

Ogden, R. 2006. Phonetics and social action in agreements and disagreements. *Journal of Pragmatics* 38(10): 1752–1775.

Pickering, L. 1999. The analysis of prosodic systems in the classroom discourse of NS and NNS teaching assistants. Unpublished doctoral dissertation, University of Florida, Gainesville, FL.

Pickering, L. 2001. The role of tone choice in improving ITA communication in the classroom. *TESOL Quarterly* 35: 233–253.

Pickering, L. 2002. Patterns if intonation in cross-cultural communication: Exchange structure in NS TA and ITA classroom interaction. *Crossroads of Language, Interaction and Culture Proceedings* 4:1–18.

Pickering, L. 2004. The structure and function of intonational paragraphs in native and nonnative speaker instructional discourse. *English for Specific Purposes* 23: 19–43.

Pickering, L. 2009. Intonation as a pragmatic resource in ELF interaction. *Intercultural Pragmatics* 6: 235–255.

Pomerantz, A. 1984. Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes. In *Structures of social action. Studies in conversational analysis*, ed. J. Maxwell Atkinson and J. Heritage, 57–101. Cambridge: Cambridge University Press.

Ramírez Verdugo, D. 2005. Nature and patterning of native and non-native intonation in the expression of certainty and uncertainty: Pragmatic effects. *Journal of Pragmatics* 37(12): 2086–2115.

Rees-Miller, J. 2000. Power, severity and context in disagreement. *Journal of Pragmatics* 32: 1087–1111.

Sacks, H. 1987. On the preference for agreement and contiguity in sequences in conversation. In *Talk and social organization*, ed. G. Button and J.R.E. Lee, 54–69. Philadelphia: Multilingual Matters.

Schegloff, E. 1998. Reflections on studying prosody in talk-in-interaction. *Language and Speech* 41: 235–263.

Selting, E. 1996. Prosody as an activity-type distinctive cue in conversation: The case of so-called 'astonished' questions in repair initiation. In *Prosody in conversation*, ed. E. Couper-Kuhlen and M. Selting, 231–270. Cambridge: Cambridge University Press.

Szczepek Reed, B. 2006. *Prosodic orientation in English conversation*. New York: Palgrave Macmillan.

Tannen, D. 1986. Cross-cultural communication. In *Handbook of discourse analysis, vol. 4, Discourse analysis in society*, ed. T. van Dijk, 203–215. London/Orlando: Academic.

Tyler, A. 1992. Discourse structure and the perception of incoherence in ITAs' spoken discourse. *TESOL Quarterly* 26: 713–729.

Tyler, A., and J. Bro. 1993. Discourse processing effort and perceptions of comprehensibility in nonnative discourse: The effect of ordering and interpretive cues revisited. *Studies in Second Language Acquisition* 15: 507–522.

Vaissiere, J. 1995. Phonetic explanations for cross-linguistic prosodic similarities. *Phonetica* 52: 123–130.

Wennerstrom, A. 2001. *The music of everyday speech*. Oxford: Oxford University Press.

Wrembel, M. 2007. Metacompetence-based approach to the teaching of L2 prosody: Practical implications. In *Non-native prosody: Phonetic description & teaching practice*, ed. J. Trouvain and U. Gut, 189–210. Berlin: Mouton de Gruyter.

Yuan, J., L. Shen, and F. Chen. 2002. The acoustic realization of anger, fear, joy and sadness in Chinese. Paper presented at the ICSLP, Denver.

Zhang, H. 2001. Culture and apology: The Hainan Island incident. *World Englishes* 20(3): 383–391.

Chapter 13
Trouble Spots in the Learning of English Intonation by Spanish Speakers. Tonality and Tonicity

Francisco Gutiérrez Díez[†]

13.1 Introduction

In the following pages I will try to describe the main difficulties to be faced by Spanish-speaking learners of English in dealing with the intonation of the latter language, more specifically, errors regarding tonality and tonicity. For the sake of transparency, I shall do it against the backdrop of the British tonetic approach, since its tonetic notation is readily interpretable and transparent in pedagogical contexts.

Three functional intonation subsystems, tonality, tonicity and tone underlie, in varying degrees of development and explicitness, several models within the tonetic approach to intonation and were explicitly labelled as such by Halliday (1967) and adopted in later models (notably Crystal 1969; Cruttenden 1986; Tench 1996, among others). I will explicitly refer to learners' errors relating to the subsystems tonality and tonicity. Tonality refers to the functional implications of the division of the speech chain into varying numbers of tone units. A tone unit is a stretch of utterance which has at least one prominent syllable (the *nucleus*) with a major pitch movement (Crystal 1969). Tonicity refers to the change of meaning brought about by placing the nucleus on alternative words within one and the same tone unit. The subsystem called tone refers to the meaningful contrasts caused by varying pitch movements, starting at the nucleus and comprising any syllable from the nucleus until the end of the tone unit. I am thus interested in the use of intonation to divide the chain of speech into information units (tonality) and to signal information focalisation within such units (tonicity). It is my contention that, strictly speaking, tonicity in its broadest sense – presence *vs*. absence of pitch accents on lexical

[†]Deceased

F. Gutiérrez Díez (✉)
English Philology Department, Faculty of Arts,
University of Murcia, Murcia, Spain
e-mail: fgut@um.es

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_13,

items – would include the focalisation of information not only by means of nucleus assignment but also by the assignment of the first accented syllable of the tone unit, generally known as the *onset,* whenever this element it is present in the tone unit, thus focalising the word containing that syllable.

Intonation errors can be interference (also called negative transfer) errors or developmental errors. The former are prompted by the transfer of mother tongue features to the students' interlanguage during the process of learning the target language. For this reason I will also point out some intonational contrasts between the two languages as a source of learning interference. Developmental errors cannot be accounted for by interference since they are produced regardless of the learner's native language and are of the same nature as those found in first language acquisition. Interference errors are not the result of developmental processes, while developmental errors stem from developmental processes which are part of universal grammar.

At this point I will explain the symbols appearing in the examples used in this paper: the symbol // stands for tone-unit boundary; an underlined word is the nucleus of the tone unit; the symbol (¯) is placed before the tone unit onset; the symbol (*) means a wrong realization of tone-unit boundary, onset or nucleus and is placed accordingly before //, an onset or a nucleus.

13.2 Tonality

In a corpus-based study of contrastive tonality (Gutiérrez 1982, 1995) among English, Spanish and Catalan the average tone unit length for the corresponding corpora used in the study was 4.36, 3.94 and 4.12 words, respectively. The "tendency towards a constant tone-unit length" (measured in number of words per tone unit) was posited as a non-linguistic variable interacting with the functional tonal delimitation/tonal cohesion variables. The chunking of utterances is determined by a trading relationship between functional pitch delimitation/pitch cohesion (i.e., the presence/absence of a tone boundary between chunks of speech) and the tendency to a constant tone-unit length (measured in number of words per tone unit). Functional pitch cohesion (the absence of // at syntactic junctions is as important as functional pitch delimitation (the presence of // at syntactic junctions; that is, the integrative aspect is as important as the delimitative one, to the extent that both of them conform to the intonational subsystem called tonality. Tonality in English has a much higher functional load than it has in Spanish, as shown by many cases where it is the only disambiguating factor in a small number of close systems; i.e., in pairs of utterances that have an identical lexical string and are distinguished solely by their intonation. Let us consider a few of these systems whose contrastive nature in the two languages can lead to (interference) tonality errors by Spanish learners of English:

(a) Relative clauses are a first case in point: other things (i.e., lexis and word order) being equal, in the example (1a–b) tonal cohesion and tonal delimitation are shown as the only means of linguistic discrimination in a close linguistic system (i.e., tonal cohesion between *bishop* and *who* in defining relative clauses (1a) *vs.* tonal delimitation at the same junction in non-defining relative clauses (1b).

(1) a. // The bishop who defied the Pope // was excommunicated // (defining relative clause)
b. // The bishop // who defied the Pope // was excommunicated // (non-defining relative clause)

Tonality in relative clauses depends on the language medium (spontaneous speech/reading aloud) and on the type of pronoun used. Except for the *zero* relative pronoun (i.e., the absence of an explicit relative pronoun) optionally used in English defining relative clauses, there is no contrast between English and Spanish regarding this type of subordinate clauses. In spontaneous speech the presence/absence of a tone-unit boundary between the relative pronoun (other than zero pronoun in English) and its antecedent have a random distribution against the variables 'defining' and 'non-defining' in both languages, whereas in read-aloud speech there is also a strong correlation between // and the syntactic junction between the relative and its antecedent for non-defining clauses in both languages and a strong correlation between tonal cohesion (absence of a tone-unit boundary) between the relative pronoun and its antecedent in defining clauses in both languages. So far, this parallelism between the two languages should not pose any learning problem for Spanish learners of English. But there is one such problem when the pronoun used in English is the zero relative pronoun, which is used with defining relative clauses as in (2a), and always requires tonal cohesion (i.e., the absence of // between the zero pronoun and its antecedent); Spanish learners of English tend to overlook that close juncture by placing // instead, as in (2b). Apparently, that is a developmental error that can only be overcome as students become familiar with the 'underlying' relative pronoun.

(2) a. // it's not a story you're gonna publish that night… //
b. // it's not a story * // you're gonna publish that night… //

(b) The intonation of final vocatives is also contrastive: in English there is tonal cohesion between the vocative and whatever precedes it, as in (3a); it can therefore be distinguished from a noun in apposition, as in (3b), where tonal delimitation is used by inserting // between the two appositional elements. Besides the vocative is either the tail of a simple nuclear tone or the rising branch of a complex fall-rise tone. Contrariwise, in Spanish there is tonal delimitation immediately before the vocative and the vocative always bears a nuclear tone. As for the distinction in Spanish between final vocatives and the second element in an apposition, it is not due to intonation, which is exactly the same in both cases, but to contextual factors. This contrast between the two languages prompts an interference error made by our learners that consists of placing // immediately before the vocative.

(3) a. // that's the technician we need // Tim//(*technician* and *Tim* are the same person)
b. // that's the technician we need Tim // (*technician* and *Tim* are different persons)

(c) In noun clauses introduced by *that* in English and by *que* in Spanish, the presence/absence of a tone unit boundary is not functional and could be

explained by a non-functional variable: the tendency to a constant length of tone units (Gutiérrez 1995). But the use of the so-called 'zero conjunction' (i.e., the omission of *that* at the syntactic junction between the main and the subordinate clauses) entails the use of tonal cohesion at such junctures. Such a construct has no parallel in Spanish and learners of English have to learn it as a specific feature of that language. During the learning process they are likely to miss the implicit underlying subordinator and, as a consequence, the compulsory absence of // at the corresponding junction, thus producing a developmental error.

(d) Another close system is represented by non-transferred negation *vs*. transferred negation (4a–b). In English, non-transferred negation is signalled by tonal delimitation between the main and the subordinate clauses (4a) and transferred negation is signalled by tonal cohesion between the main and the subordinate clauses (4b). In Spanish, the same distinction is signalled by the use of the indicative mood for non-transferred negation (5a) and the subjunctive for transferred negation (5b–c).

(4) a. // I didn't stay/because it was raining // ('I didn't stay')
b. // I didn't stay because it was raining // ('I stayed for some other reason')

(5) a. // no me quedé // porque estaba lloviendo //
b. // no me quedé porque estuviera lloviendo //
c. // no me quedé // porque estuviera lloviendo //

In cases like (4b), Spanish learners usually miss the compulsory tonal cohesion, which can be explained by the fact that in Spanish, tonal delimitation/tonal cohesion at the same categorical junction (5b–c) are not functional and, therefore, are randomly distributed at the junction between the main and the subordinate clauses.

In the corpora I have used so far, a common feature is the excessive number of tone units produced by Spanish learners of English in comparison with native speakers of English. In one of them (Gutiérrez 2005), the reading aloud of an English text by both native English speakers and high school Spanish learners of English in their second year of Bachillerato (sixth-formers) yielded the following results: the Spanish informants produced about five times the number of tone units produced by the English ones (an average proportion of 185 to 40 tone units, respectively). A somewhat similar result is reported in a study by Gutiérrez (2008). In this case, 15 Spaniards who were learners of English in their fourth year of ESO (Secondary Education) produced an average of 70 tone units per informant, as compared to the 40 tone units produced by one English native informant.

The explanation should probably be sought in the students' limited command not only of the segmental content of pronunciation, but also of the morphosyntactic, semantic and pragmatic levels and, as a consequence, in their limited prosodic fluency. If this were the case, and in view of the similar tone-unit length referred to above for Spanish and English by native speakers, it seems plausible to label

most tonality errors (among them the excessive fragmentation of the speech chain into tone units) as developmental. As such they can be expected to naturally disappear from the learners' interlanguage as they gain an increasing command of both segmental and rhythmic pronunciation and improvement in their handling of morphosyntactic aspects, thus leading to greater prosodic fluency in the use of language.

As we shall see below, this excessive number of tone units produced by the learners may explain, at least in part, what goes on in the rendering of English tonicity by the same type of learners: namely an excessive number of nuclei and onsets in comparison to the output of English native speakers for one and the same speech sample. A striking type of developmental tonality error consists in placing tone-unit boundaries in the wrong syntactic junctions, as is the case in (6), where the tone unit boundary between 'always' and 'had' is quite unnatural, since English native speakers will not expect a boundary between a main verb and its auxiliary because it contributes to distorting perception.

(6) … // as ¯Tom Cruise // *¯has always // had an easy life // but ¯that isn't true //

From a categorical perspective this type of error is also made by Spanish children learning their native language; hence, the developmental character of such errors. It can also be safely termed a phonetic error, whereas errors related to the functional uses illustrated in (a–c) above are phonological errors. The difference is that the former are a sign of distorted fluency and a feature of foreignness, without their hindering the perception of syntactic categories and/or the ensuing pragmatic meaning by listeners, whereas the latter, in the absence of other contextual cues, may cause a breakdown in communication. It is my hypothesis that phonetic errors of this type do not need special pedagogical treatment since they will take care of themselves as the student progresses towards intermediate and advanced learning stages. This is in sharp contrast with what happens with segmental phonetic errors, which can become fossilised, as is the case, for example, with many advanced Spanish speakers of English who substitute a voiced velar fricative [ɣ] for a voiced velar plosive [g] in intervocalic position (as in the word *again*).

13.3 Tonicity

The prosodic prominence given to certain words in the speech chain aims at focalising the information content of such words and it is formally assigned by means of a combination of rhythmic stress and pitch accent. Traditionally, lexical stress has been considered a lexical phonological feature. It serves as input to rhythmic stress, whose rules may determine the redistribution of lexical stress within certain words of the speech chain (like *princess, canteen, seventeen,* etc.) and its suppression in others, as, for example, in function words, which typically occur in the unstressed

part of rhythmic units. The output of rhythmic rules becomes an input to pitch-accentuation (i.e., intonation) rules. The latter yield the final configuration of pitch prominence within and between intonation units.

Though traditionally as a means of focalising information, tonicity has been linked exclusively to the assignment of the intonational nucleus, it is my opinion that since onset assignment is another way of highlighting information by singling out the onset-bearing word by means of pitch accentuation, perhaps it is both theoretically and pedagogically sound to consider onset placement as an "extended" way of tonicity, even if the information focalisation taking place at onset position is arguably far less important than the one carried out at nucleus position. That is why I shall include errors related to onset placement at this point in the present paper.

13.3.1 Onset Assignment

The onset syllable is simultaneously the first rhythmically-stressed syllable and the first pitch-accented syllable of the tone unit; it is an essential structural point in the shaping of heads (O'Connor and Arnold 1973) or pre-tonic segments (Halliday 1967) and of an intonation subsystem called *key,* which refers to contrasts based on the different fundamental frequencies (F_0) of the overall contours of adjacent tone units within a tone-unit sequence (Brazil et al. 1980; Couper-Kuhlen 1986).

There is a strong parallelism in the phonetic realisation of the onset in both English and Spanish: it falls on the first syllable of the tone unit with a rhythmic beat that has also a pitch accent. So far, no difficulties are foreseeable for Spanish learners of English. Problems, though, begin when they have to align the onset position with word categories and they have to decide which English words are eligible for receiving both rhythmic stress and pitch-accent in onset position. And it is with regard to such word categories that the two languages are in contrast with one another, thus yielding trouble spots for learners.

Some word categories are not eligible for occupying onset position either in English or in Spanish. Such is the case with conjunctions, possessive determiners and prepositions, which do not carry either rhythmic stress or intonational onset for emphatic reasons in either of the two languages. Yet onset errors are found in the oral output of Spanish learners consisting of the accentuation of such word categories. The symbol (¯) indicates an onset, and (*¯) indicates a wrong realisation of onset; for instance, the conjunction *that* in (6), the possessive *his* in (7) and the prepositions *because* and *as* in (8).

(6) …// was more intelligent // than he was // and *¯ that he could never succeed //
(7) // Then // in *¯his last year // of ¯secondary school //…
(8) // Be¯cause he succeeded / *¯as an actor //…

Since both languages behave in the same way as regards those three word categories, there is no question of interference errors in (6–8). Therefore, they can be safely labelled as developmental errors.

The intonational behaviour of auxiliary verbs and personal pronouns is different in the two languages involved. In Spanish they are realised with a rhythmic stress and a full vowel (i.e., a peripheral and stressed vowel). They are also susceptible to receiving an onset when they are the first rhythmically-stressed word in the tone unit. In English, on the contrary, these two types of words are weak (i.e., they are realised with a weak vowel), do not have rhythmic stress and, therefore, may not be pitch-accented, and therefore may not become tone-unit onsets. Example (9) illustrates the erroneous choice of onset on the personal pronoun *you,* and in example (10), the wrong choice is on the auxiliary *has*. Since personal pronouns and auxiliary verbs may become onsets in Spanish but not in English, we would have two (negative) transfer errors here.

(9) // ¯Tom says // ¯don't expect people to hand // *¯you things //
(10) …// for a long//long time//that I ¯felt excited // about // he *¯has often said //

13.3.2 Nucleus Assignment

Nowadays, we are in a position to correct Halliday's maximalist position according to which the meaning of intonation is grammatical (Halliday 1967, 1970). Tonicity is a good example of how intonation functions in the pragmatic arena. Notions such as shared/non-shared information, new/given information can only be generated in actual texts and are not directly associated to intonational forms out of texts. Those notions emerge as pragmatic meanings from the mutual interaction of parts of one and the same tone unit or from the mutual interaction of different tone units in the text. Only then can they be associated with intonational forms (i.e., presence/absence of an intonational nucleus).

In this section I will try to show that the fact that both neutral tonicity and marked tonicity are contrastively patterned in the two languages, leads to the systematic appearance of interference errors in the tonicity subcomponent of our learners' interlanguage.

Neutral tonicity (Halliday 1967) means the assignment of the tone-unit nucleus to the final lexical (i.e., non-grammatical) word of the tone unit. There are sharp tonicity contrasts between English and Spanish regarding the alignment of nucleus with function words. Such an alignment occurs systematically in Spanish when the function word occurs in tone-unit final position. However, in English, function words in final position may not be nucleus bearers, except in cases of marked contrastive focus (marked tonicity). In Spanish, neutral tonicity would consist in the assignment of the nucleus to the final word (whether lexical or grammatical) of the tone unit. Regarding articles, prepositions and conjunctions, there is parallelism between the two languages in so far as these words are not candidates either for rhythmic stress or nucleus assignment; consequently, tonicity errors consisting of nucleus assignment to such word types are developmental errors. An example is the assignment of nucleus to the article *a* in (11). Another one is the assignment

of the nucleus to the conjunction *since* in (12). Nucleus placement on English personal pronouns (*him*, *them*, *it*, *us*, etc.) by the learners constitutes a type of error that can be classed as interference error, as is the case with *him* in (13). Learners will have to learn not to place the nucleus on English personal pronouns whether they have subjective or objective form. In Spanish, the norm is different and twofold: in tone-unit final position non-enclitic pronouns are stressed, but enclitic ones (*-te, -le, -las, -nos, -os,* etc.) are unstressed, probably because they are not phonologically independent words, since they are phonologically integrated with the word preceding them. This would explain cases of interference when learners tackle the prosody of English pronouns. The most frequent error of this type is made with personal pronouns in final tone-unit position (Gutiérrez 2005).

(11) // He has become *a // wealthy man // but more important for Tom // he has learnt to read //
(12) …// everywhere *since // he was // nineteen years old //
(13) // He was always // in trouble // and felt // that no one // understood *him //

Marked tonicity consists in the assignment of a nucleus to a lexical word other than the last lexical one because it carries new information (i.e., it has become the information focus). From a pedagogical point of view, marked tonicity implies that the learner has to learn to de-accent words which carry old or given information, starting from right to left in the tone unit, until reaching a new-information-bearing word on which to place the nucleus. Marked tonicity also serves the purpose of signalling contrastive information. In example (14) the nucleus should have been placed on the contrastive element *he,* rather than on the item *was*, for two reasons: because *he* contains contrastive information (it contrasts with *everyone*) and because the element *was* contains given information. In this case, we also have an interference error. In (15) the word *thing* carries given or familiar information, which should have prompted the shifting of the nucleus to the word *first*, which is now the carrier of new information.

(14) // Everyone was more intelligent than he *was //…
(15) // it was the first *thing //… // that I felt excited *about //

In the previous paragraphs it has been shown that not only marked tonicity but also neutral tonicity is contrastively patterned in the two languages. Such contrasts, as we shall see below, constitute a stumbling block for our learners, who make not only developmental errors, especially at the basic level of English, but also interference errors, which easily become fossilised and carried over during the advanced level, as pointed out by Ramírez (2003: 662) regarding a group of university students, who besides and before the 3-year treatment pointed out by the author, had been exposed to English for 6 years during their secondary education.

> The longitudinal study demonstrates that there has not been any progress in the production of English intonation by Spanish learners…The results prove that the acquisition of intonation does not take place automatically, even though learners have been exposed to English for about three years. (2003: 662).

Table 13.1 Distribution of tone units against the variables neutral *vs*. marked tonicity in English and Spanish

	Neutral tonicity	Marked tonicity	Number of tone units
English	1,294	77	1,371
Spanish	1,497	9	1,506

Research has shown that, although marked tonicity is not completely alien to Spanish (García Lecumberri 1995), it is scarcely used, especially at the production level. A clearly contrastive feature is that de-accenting old information is compulsory in English, whereas in Spanish it is neither compulsory nor very frequent.

Ortiz Lira (1994) found nuclei followed by de-accented material in only 8–10% of all the intonation units in his Spanish corpus of data. In Table 13.1 results are shown concerning marked tonicity in half an hour of televised panel discussion by native speakers of English and another half an hour for the same type of speech by native speakers of Spanish (Gutiérrez 1995).

According to the results in Table 13.1, marked tonicity was used in only about 0.60% of the Spanish tone units. Cruttenden (1993) finds that some languages, amongst them English, insist on de-accenting repeated material, while others, like Spanish, strongly resist it. In line with the foregoing, García Lecumberri (2000) reports on the poor performance of Spanish learners of English at perceiving the informative focus at nucleus position, especially when the utterances contained a marked focus.

The higher functional load of tonicity in English as compared to Spanish is evident in cases of utterances with an identical lexical string in which tonicity is the exponent of two different meanings. That is what happens, for instance, when in a pre-modifying structure, tonicity may alternatively signal a determiner *vs*. an adverbial intensifier. In (16a) *more* is a quantifying determiner ([*more* [*effective teachers*]]) and receives the nucleus; in (16b) the same word is an intensifier ([[*more effective*] *teachers*])

(16) a. // we need <u>more</u> effective teachers //.
b. // we need more e<u>ffec</u>tive teachers //

In Spanish, such contrasts are not signalled by intonation but by word order ((a) *más profesores efectivos* vs. (b) *profesores más efectivos*). This is one of those specific uses of English intonation that represent an additional burden for Spanish learners of that language and needs specific treatment by teachers during the learning process if fossilisation of related errors is to be avoided.

As I said before, many tonicity errors are prompted by concomitant tonality errors, which act as input to the former. More specifically, developmental tonality errors prompt the emergence of interference tonicity errors.

Most onset and tonicity errors found in our corpora are related to ontogenically previous rhythmic errors, such as stressing function words in positions and under conditions in which English would not allow rhythmic stress.

It has been shown that some tonicity errors can only be fully accounted for by reference to the learners' wrong performance regarding tonality: the wrong insertion of an excessive number of tone-unit boundaries leads to an excessive number of nuclei and nuclear tones. This supports the widely-held theoretical view according to which the three intonational subsystems of tonicity, tonality and tone are tightly interrelated at both competence and performance levels.

13.4 Concluding Remarks

It is my contention that in a learning context that takes place away from an English speaking country, intonation cannot be acquired unless it is explicitly taught.

Most developmental tonality errors do not need any special care, since they will disappear by the end of the intermediate stage as fluency at the segmental level (fluent use of sounds in words, phrases and sentences) increases.

As for teaching students to avoid onset misplacement, I would like to suggest that the best strategy is to teach them to correctly place rhythmic stress. To the extent that they learn not to stress English grammatical words, onset placement will cease to be a learning problem and so would be instances of neutral tonicity where grammatical words appear in tone-unit final position. The learning of onset placement and of neutral tonicity is a matter of learning not to stress grammatical words but lexical ones. The type of onset to be used (high, low, mid, etc.) – a matter not included in the present study – would also have to be explicitly taught in connection with the intonational subsystem called *key,* which is realised between sequences of tone units by, among other things, allocating high, mid or low levels to the onsets of each of the tone units making up such sequences.

Interference tonicity errors, such as nucleus placement on (non-rhythmically stressable) grammatical words can only be overcome by explicit pedagogical attention at the early learning stages. Otherwise, error fossilisation will crop up. Given the strong contrastive nature of tonicity in the two languages involved, it is no surprise that marked tonicity in English is by far the most stubborn stumbling block of all three intonational subsystems for Spanish speakers learning English. It needs a good amount of explicit teaching. This must include training Spanish students to read the pragmatic meanings 'newness/givenness' in real texts as determined by the mutual interaction of textual segments; and teachers can do this by using simple, comprehensible language such as 'this item is being mentioned for the first time in the text and that is why it has a nucleus' (accenting new information) or 'this word has already been mentioned in the text and that is why it does not have a nucleus' (de-accenting information). Unless learners get used to recognising the coupling of newness/givenness with the presence/absence of a nucleus by conscious practise of that association in real texts, they will not be able to produce such couplings

themselves (also through practice). The alternative is the likely fossilisation of marked tonicity errors. Once the students become familiar with the two most frequent patterns of marked tonicity mentioned above, they can be exposed to the exception represented by accenting given/contrastive information, as triggered by the need to *emphasise* such information or to *contrast* it with previous information.

A good diagnosis of some of the reasons for deficiencies in the pronunciation of our students of primary and secondary education in Spain can be found in Jódar (2005) and Martínez (2004). Rhythm and intonation have traditionally been the 'Cinderellas' of the pronunciation component in the classroom. The official inclusion of oral skills and pronunciation as test targets in the university entrance examination by the academic year 2011–2012 seems promising, since that decision will no doubt prompt the implementation of that component in textbooks and in real teaching practice.

Further studies of tonicity errors at the learning levels of primary, secondary and university education would provide a better view on the profile and development of intonation errors. Further quantitative studies are needed from both cross-sectional and longitudinal perspectives, in order to gain a wider and richer picture of the combination of three variables: types of errors, learning rates and error distribution along different proficiency levels.

References

Brazil, David, Malcolm Coulthard, and Catherine Johns. 1980. *Discourse intonation and language teaching*. London: Longman.

Couper-Kuhlen, Elisabeth. 1986. *An introduction to English prosody*. London: Edward Arnold.

Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.

Cruttenden, Alan. 1993. The de-accenting and re-accenting of repeated lexical items. Working Papers, Department of Linguistics and Phonetics, Lund [*Proceedings of the ESKA Workshop on Prosody*, Lund], 41:16–19.

Crystal, David. 1969. *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

García Lecumberri, M. Luisa. 1995. *Intonational signalling of information structure in English and Spanish.* Bilbao: Servicio de Publicaciones de la Universidad del País Vasco.

Gutiérrez, Francisco. 1982. Bases para una comparación de la prosodia del catalán, elcastellano y el inglés. Dissertation, University de Barcelona.

Gutiérrez, Francisco. 1995. *La función demarcativa de la entonación en inglés castellano y catalán*. Murcia: Servicio de Publicaciones de la Universidad de Murcia.

Gutiérrez, Francisco. 2005. Intonation focus in the interlanguage of a group of Spanish learners of English. *Revista Alicantina de Estudios Ingleses* 18: 127–147.

Gutiérrez, Francisco. 2008. Contrastive intonation and error analysis. Tonality and tonicity in the interlanguage of a group of Spanish learners of English. In *Languages and cultures in contrast and comparison*, ed. M.A. Gómez, J. Lachlan Mackenzie, and E. González, 327–354. Amsterdam: John Benjamins.

Halliday, Michael. 1967. *Intonation and grammar in British English*. The Hague: Mouton.

Halliday, Michael. 1970. *A course in spoken English*. Oxford: Oxford University Press.

Jódar, Oscar. 2005. Estudio sobre los efectos de un tratamiento alternativo del componente fonológico en el currículo de lengua inglesa para 2º de Bachillerato. Unpublished Ph.D. thesis, University of Murcia.

Luisa García Lecumberri, M.Luisa. 2000. Identificación del foco entonativo inglés por estudiantes frente a hablantes nativos. *Estudios de Lingüística Aplicada* 32: 19–28.
Martínez, Francisco. 2004. Estudio de una intervención pedagógica para la enseñanza de la pronunciación inglesa en 4º de la ESO. Unpublished Ph.D. thesis, University of Murcia.
O'Connor, Joseph, and Gordon F. Arnold. 1973. *Intonation of colloquial English* (1st edn 1961). London: Longman.
Ortiz Lira, Héctor. 1994. Contrastive analysis of English and Spanish sentence accentuation. Dissertation, University of Manchester.
Ramírez, M.D. 2003. A contrastive analysis of non-native interlanguage English intonation systems and their implication in the organization of information from a functional perspective. Ph.D. dissertation, Universidad Autónoma de Madrid.
Tench, Paul. 1996. *The intonation systems of English*. London: Cassell.

Chapter 14
Teaching Prosody with a Pragmatic Orientation: A Synthesis

Jesús Romero-Trillo

The chapters of the present volume have shed light on the multifaceted aspects of English prosody and on the transmission of meaning in the language teaching context. If the analysis of prosody and pragmatics in isolation already poses a good amount of theoretical and practical problems for linguists, their factual liaison for English language teaching demands deeper reflections on how the language class can be an optimal laboratory, i.e., an autonomous place with controlled practice that resembles and fosters the use of language in real contexts. The aim of the volume is to describe the creation of meaning through the conjunction of linguistic and extralinguistic features, such as gaze, body posture, etc. The objective of this chapter is to suggest strategies that teachers might find useful for their daily practice.

Firstly, language teachers need to tackle the teaching of prosody and pragmatics in unison, with a clear idea of the first language features of their students prosody-wise. As the previous chapters have explained, one of the most important concepts is that there are two categories of languages: tone languages in which lexical elements differ semantically in terms of tone variation (lexical tones), like Chinese for example; and non-tone languages, like English. In tone languages lexical distinctions are realized through tone variation, whereas in English prosody realizes discourse/pragmatic functions: to convey the focus of information, to encode the sentence type, to initiate turns/topics, to indicate contradiction and other stylistic phenomena, to avoid misunderstanding, etc. In other words, teachers need to know the type and functions of the prosody of their student's first language. In the case of English, prosody has a discourse/pragmatic orientation that helps the speakers identify or infer information that relates not only to the actual meaning of the words, but also to the pragmatic and discourse organization of speech. The specific discourse orientation of English prosody may foster the contextualized practice of prosodic

J. Romero-Trillo (✉)
Departamento de Filología Inglesa, Facultad de Filosofía,
Universidad Autónoma de Madrid, Madrid, Spain
e-mail: jesus.romero@uam.es

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*,
Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6_14,

features, above the sentence borderline if possible, in order to help students perceive the full-fledged pragmatic hues of English prosody. To achieve this, students will have to be exposed to, and produce, pragmatic meanings at the discourse level (focus assignment, disagreement, etc…) in which very similar utterances might change their pragmatic orientation thanks to the specific prosodic realization.

One important fact to remember at this point is that babies experience prosody before birth. Therefore, all students are perfectly equipped to imitate prosodic features, sometimes with even more success than with the pronunciation of phonemes that are alien to their first language. In fact, there are certain features that are deemed universal, like the use of rising tones for questions although, obviously, others are language specific. The awareness of the prosodic features that are shared with English is essential as students can hinge upon them at least in some basic utterances like statements or questions. However, the universal presence of prosody in all languages is sometimes a drawback for its correct study in a second language and, as teachers and students know, English can be especially problematic in this respect. Recent research shows that English presents different ways of expressing functions, although without a one-to-one correspondence between form and function, which makes foreign speakers makes foreign speakers strive to create meaning in context with prosodic accuracy at the discourse level. As teachers know, it is not easy to reach a native-like prosodic performance in English and some studies have shown the relationship with Length of Residence and Age of Arrival, as two objective quantitative variables that can play a role in second language prosodic development. In fact, it has been demonstrated that the attitude of native speakers towards foreign speakers, which can vary a lot in terms of social and first language origin, may also influence performance.

One of the elements that are first perceived in the comparison between native and non-native speakers of English is stress. Stress can be described as the pace of recurrent beats that mark the rhythm of the language. Languages have been traditionally classified as stress-timed (those whose recurrent beats has a regular pattern), like English, and syllable-timed languages (whose basic measurement unit is the syllable), like Spanish or German amongst many others. Recent approaches to rhythm have modified this conception and now distinguish between stress- and syllable-based languages. Again, English language teachers should identify the category their students' first language belongs to and act accordingly in terms of stress practice, after considering the specificity of English rhythm.

However, not all 'Englishes' have the same rhythmic pattern. Scholars differentiate between 'Inner-Circle Englishes', with stress-based rhythm, and 'Outer-Circle Englishes', with syllable-based rhythm. Inner-Circle English coincides with the countries where English has been the native language for centuries (UK, Ireland, USA, Canada, Australia, etc.), while the latter variety is found in the countries where English has been adopted as a (second) language for daily communication (African and Asian countries mostly). In fact, many foreign learners whose first language is syllable-based will find these 'Englishes' much easier for communication, as words are spelt out with 'more clarity' to their ears in comparison with 'Inner-Circle Englishes', in which some words – especially grammatical items – are reduced and sometimes even inaudible for untrained speakers. Making students

aware of this fundamental, though often neglected, feature can result in a dramatic improvement in their comprehension and communication skills, as they will be able to identify informative segments that were originally imperceptible for them.

From a cognitive perspective, some scholars have suggested that speakers who use stress-based languages might process information differently from those with syllable-based languages, as the salient prosodic information is encoded differently and, therefore, their cognitive focus would be processing the message diversely. Although there is still debate on this hypothesis, it is clear that the pragmatic implications conveyed through the stressing of grammatical items – pronouns, auxiliary verbs, prepositions etc. – by non-native speakers would highlight elements that are not typically stressed by native speakers, except for emphasis or contrast, which may trigger some sort of puzzlement in the listeners. Careful training on the pronunciation of unstressed grammatical items is a very important task in ELT, with the subsequent teaching of the pragmatic and discursive role of stressed grammatical items for emphasis or contrast.

An accurate prosodic choice is a crucial element for the creation of context, and its awareness is especially important for non-native speakers of English. Prosody is a dynamic element moulded along the development of a conversation with the interplay of given and new information through tonicity. The mismatch of grammatical and prosodic choices in the expression of new information results in anomalous prosodic patterns that not only betray the non-nativeness of the speaker, but also the indefiniteness or ambivalence of the information focus. Other pragmatic functions, such as communicative intent, or the expression of emotions and attitudes, are also conveyed through prosody, which does weigh more than the actual meaning of the words in many cases; e.g., if someone says 'I feel happy' with a low tone and slow tempo, the listener will possibly understand the opposite. In this sense, humour, irony and other related frequent elements of communication rely on the synergy of prosody with lexis and grammar. This is a crucial element to practice with learners of English, to provide them with the linguistic tools to interact in real contexts.

However, the use of prosody in the creation of context is not only related to 'creativity' but also to pragmatic aspects, such as to avoid misunderstanding. For example, Pragmatic Markers – 'you know', 'I mean', 'well', etc. – are used to guarantee the communication process from a discourse, pragmatic and cognitive perspective. These elements are especially interesting because they can appear in conversation with multiple prosodic patterns depending on the pragmatic intent of the speaker and on the reaction of the listener. This is the reason why the erroneous use of these elements to start a turn, initiate a topic, offer feedback or correction, might lead to pragmatic misunderstanding and, if uncorrected, to pragmatic fossilization. Learners of English must pay special attention to the use of these elements and also be aware of their multiple prosodies in comparison with their first languages, because the variety, frequency and distribution of Pragmatic Markers is language specific and not immediately transferrable, especially between English and other languages.

In practical terms, the use of computer programmes, such as 'Praat' or 'Speech Analyzer', allows teachers – and also advanced learners – to monitor their performance with easy-to-interpret graphs on the computer screen. Students can record

their speech and compare it with model sequences previously uploaded by their teachers. These recordings can start with isolated utterances focusing on features of stress, rhythm or intonation. Nevertheless, it is very important that contextualized chunks be analysed for pragmatic features, with comparable examples in which students can identify meanings only realized through prosody. The repetition and analysis of these sequences is fundamental for the proper development of pragmatic competence.

As in all educational practices, the role of the teachers is essential and their knowledge of English prosody and the learners' first language prosody could be a crucial asset. The teachers' practical performance in class with a careful production of prosodic patterns can also often illustrate English prosody better than textbooks. For instance, some studies have shown that native teachers have a wider choice of tone sequences than non-natives, which results in a higher exposure to a greater variety of tones for different pragmatic meanings. The awareness of the multifaceted correspondence between tones and meanings in context will help non-native teachers increase their prosodic competence and have a more native-like performance. For instance, this relationship is especially noticeable in pitch concord and in the capacity to liaise pragmatic meanings in turn-taking. A careful practice for basic discourse functions like agreeing – expressed with pitch concord – and disagreeing – expressed without pitch concord – can help learners feel more at ease in casual conversation, as they will be attentive not only to the words expressed by the interlocutor to show (dis)agreement, but also to the pitch concord which, sometimes, could be the determinant for correct understanding.

Another clear example is the teaching of tonality – the division of speech in meaningful segments – as for instance in the practice of relative clauses (finite and non-finite) and for final vocatives (often used in classroom context to draw the students' attention). A meticulous division of speech into tone units will make learners identify prosodic chunks with units of meaning, which are marked with the tonic elements that weave the meaning progression through the interaction.

As this volume has shown, the students' awareness of how meaning is paved with prosodic choices that constitute cognitive and tonic stepping-stones is vital for their linguistic competence. In fact, in order to guarantee communication in real contexts, students and teachers have to implement the – often – unattended synergy of pragmatics and prosody. Both disciplines face communicative challenges from different but complementary perspectives: the acoustics and intentions in a given context. As all speakers know, judging someone's intentions by just the words without paying heed to his or her 'way of saying' is impossible or daring in real life. Likewise, guessing someone's intentions just by prosody without considering the exact words and context would be unwise. To sum up, this volume has delved into the symbiosis of pragmatics and prosody within the English teaching context and has shown the challenges and possibilities of applying current theoretical advances to the improvement of learners' communicative skills.

Author Index

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6,

Subject Index

J. Romero-Trillo (ed.), *Pragmatics and Prosody in English Language Teaching*, Educational Linguistics 15, DOI 10.1007/978-94-007-3883-6,